# Deterministic and Probabilistic Calculations to Estimate Risk-Based Cleanup Goals for Soils at Residences Near the 2800 South Sacramento Site, Chicago, Illinois

25 October 1996

Alceon Corporation
PO Box 382669
Harvard Square Station
Cambridge, MA 02238-2669

617-864-4300

# Executive Summary

## Introduction/Purpose of Report

Alceon Corporation has prepared this risk assessment to estimate Risk-Based Cleanup Goals (RBCGs) for soils at residences near the 2800 S. Sacramento site in Chicago, Illinois. The Celotex Corporation, AlliedSignal, and the US Environmental Protection Agency Region V (US EPA) have had several discussions about draft risk assessments previously performed for this site. Although Celotex, AlliedSignal and US EPA have agreed on some exposure assumptions, they differ on others. Therefore, the purpose of this report is fourfold:

1. To develop RBCGs for benzo(a)pyrene equivalents (BaPeq) deterministically using exposure assumptions proposed by Celotex and AlliedSignal.

2. To develop RBCGs for BaPeq as a distribution, using probabilistic methods.

3. To present, side-by-side, the RBCGs proposed by Celotex, AlliedSignal, and the US EPA Region V to facilitate understanding of the differences.

4. To present the outline of a health-protective, resource-efficient method for selecting residential properties for remediation should the need be demonstrated in the risk assessment.

## Site Location and Description

The 2800 S. Sacramento site is located at 2800 South Sacramento Avenue in Chicago, Illinois. The site, including the 18-acre Celotex property and a 6-acre property to the south (which is currently being used for truck storage), is located in a mixed industrial, commercial, and residential area. Industrial buildings on the Celotex property were demolished and removed in 1993, except for some concrete slabs and foundation remnants, and a soil cover was then placed on the property. The entire site is surrounded by a chain-link fence.

## Site History

Before 1912, the area was a farmstead. Allied Chemical Corporation operated a coal tar distillation and roofing plant at the facility beginning in 1912. The facility was sold to The Celotex Corporation through several transactions between 1967 and 1979.

## Regulatory Background

The Illinois Environmental Protection Agency (IEPA) conducted various investigations at the 2800 S. Sacramento site from 1989 to 1993. In 1993, a US EPA Technical Assistance Team assessed the area, and in 1994, the US EPA issued a Special Notice of Liability and proposed Consent Order to Celotex and AlliedSignal. The proposed Consent Order required Celotex and AlliedSignal to perform sampling in residential areas in the vicinity of the site, prepare an Engineering Evaluation/ Cost

Analysis (EE/CA), and perform remediation as appropriate. AlliedSignal and Celotex signed an informal agreement with the US EPA in July 1995 under which the companies performed an investigation of contamination on residential properties within the study area, which was defined as the homes within a 2,500-ft radius from the site.

## Chemicals of Concern

The carcinogenic polycyclic aromatic hydrocarbons (PAHs) selected as the study chemicals are expressed as BaPeq; these chemicals were selected because they dominate the calculations of RBCGs. PAHs generally adsorb strongly to soils, migrate slowly if at all in ground water, and do not readily volatilize into soil gas or the atmosphere.

## Methods Used in This Risk Assessment

Alceon has used both deterministic and probabilistic calculations to estimate RBCGs for surface soils in the study area. The risk assessment is based on three pathways, incidental soil ingestion, inhalation of fugitive dust, and dermal contact with soil. The RBCG based on exposure assumptions proposed by Celotex and AlliedSignal for Reasonable Maximum Exposure (RME) was calculated first using deterministic methods. Then the RBCG as a distribution was calculated using probabilistic methods. The last calculation was based on RME assumptions proposed by US EPA. The only difference between the companies' calculated RBCGs and the US EPA's is in their respective exposure assumptions. All the calculations use the same toxic potencies listed in the US EPA's Integrated Risk Information System (IRIS) database.

The two sets of deterministic exposure assumptions differ because the companies rely primarily on the latest US EPA "Guidelines for Exposure Assessment" to develop RME assumptions, using information appropriate for the study area; US EPA relies primarily on earlier Agency guidance to develop default RME assumptions using "bounding estimates." Bounding estimates may be regarded as maximally conservative estimates, which, given the compounding effect that occurs in the calculations, exceed reasonable estimates of actual exposure, even for the most exposed individuals. The assumptions used by the companies for both their deterministic and probabilistic risk assessments are intended to reflect a reasonable, conservative estimate of risk.

As an example, the companies and US EPA agree that if the temperature is below freezing no children will play outdoors and have exposure to the soil. However, Celotex and AlliedSignal proposed that as the temperature rises, increasing percentages of children should be assumed to have exposure. US EPA assumed that if the temperature is above 32 degF, all children are exposed to the outdoor soil every day.

## RBCGs for Study Area

Using deterministic methods, Alceon estimated the RBCGs for surface soils at residential houses near the 2800 S. Sacramento site as follows:

| | |
|---|---|
| Celotex and AlliedSignal's exposure assumptions: | 27.5 mg/kg BaPeq |
| US EPA's exposure assumptions: | 1.93 mg/kg BaPeq |

Using probabilistic methods, Alceon calculated a distribution of BaPeq concentrations that represents the response goal, or cleanup target, for BaPeq concentration in soils. The exposure assumptions on which the calculations are based are similar to those used by Celotex and AlliedSignal for their deterministic risk assessment, this time expressed as distributions. The health-protective distribution of response goals is shown below:

| Percentile | | Cleanup Goal | Comment |
|---|---|---|---|
| minimum | = | 0 mg/kg BaPeq | US EPA's 1.93 mg/kg cleanup goal lies near extreme end of health protective distribution |
| 10th percentile | ≤ | 6.4 mg/kg BaPeq | |
| 20th percentile | ≤ | 9.5 mg/kg BaPeq | |
| 30th percentile | ≤ | 12.8 mg/kg BaPeq | |
| 40th percentile | ≤ | 16.4 mg/kg BaPeq | |
| 50th percentile | ≤ | 20.8 mg/kg BaPeq | |
| 60th percentile | ≤ | 26.1 mg/kg BaPeq | Companies' 27.5 mg/kg goal is near mid-range. |
| 70th percentile | ≤ | 33.3 mg/kg BaPeq | |
| 80th percentile | ≤ | 43.6 mg/kg BaPeq | |
| 90th percentile | ≤ | 58.2 mg/kg BaPeq | |
| 95th percentile | ≤ | 72.7 mg/kg BaPeq | |
| maximum | ≤ | 99.9 mg/kg BaPeq | |

## Identifying Potential Candidates for Remediation

Implementation of response goals in a residential neighborhood presents practical issues regarding health protection and feasibility. Part of the solution requires an understanding that individuals do not have all their exposure on a single property, but over many properties within their neighborhood, as described in US EPA guidance documents. A resulting "exposure point concentration" (EPC) can be calculated using spatial statistics based on activity-, time-, and distance-weighted factors.

The cleanup target distribution listed above provides the basis for a remediation strategy that is health protective yet resource-efficient. The approach would work as follows:

1. Characterize through sampling and analysis the BaPeq concentrations at all the residences in the study area. A sampling program may be possible that yields more information than one in which every home is sampled, since measurements within a property are subject to variability.

2. Using spatial statistics, estimate the distribution of EPCs.

3. Compare the distribution of EPCs based on measured concentrations to the distribution for the RBCG. If any percentiles of the EPC distribution exceed the corresponding percentile of the RBCG distribution, remediate the surface soils at one or more of the most contaminated properties.

4. If the distribution of EPCs based on the remaining concentrations still does not meet the distribution for the RBCG, remediate one or more of the more contaminated properties that remain.

5. Continue remediation until the distribution of EPCs is smaller (i.e. has lower percentiles) than the RBCG distribution.

This approach provides appropriate protection and minimizes neighborhood disruption, which is also a concern in the development of a remedial plan.

# Table of Contents

Map
Figures
Tables
Appendix A - Background Concentrations of Benzo(a)pyrene
Appendix B - Estimating the Spatial Extent of Site-Related Contamination (Cox Associates)
Appendix C - Toxicity Profile for Benzo(a)pyrene from US EPA's IRIS Database
Appendix D - Outdoor Exposure Frequencies for the Neighborhoods Near the Industrial Property
Appendix E - Exposure Duration for the Neighborhoods Near the Industrial Property
Appendix F - Transfer Coefficient from Soils Outdoors to Dust Indoors
Appendix G - Absorption Adjustment Factors (Oral and Dermal) for PAHs (Ogden Environmental)
Appendix H - Uncertainty Analysis of Risk Estimates (Cox Associates)
Appendix I - Copies of Peer-Reviewed Publications

# 1. Introduction

This human health risk assessment uses both deterministic and probabilistic calculations to estimate risk-based cleanup goals for surface soils at residential properties near the industrial site in Chicago, IL. Alceon has completed this human health risk assessment report in accordance with the most current quantitative human health risk assessment methods of the US Environmental Protection Agency (US EPA, 1992, 1993, 1995; US EPA, 1994, MC; Browner, 1995; US EPA, 1995, TG).

The body of this report contains both deterministic and probabilistic calculations for "risk-based cleanup goals" (RBCGs) for the concentration of total carcinogenic polycyclic aromatic hydrocarbons (cPAHs) as benzo(a)pyrene (BaP) equivalents in surface soils. First, we calculate the RBCG as a single concentration of BaPeq based on deterministic "Reasonable Maximum Exposure" (RME) exposure assumptions proposed by Celotex Corporation and AlliedSignal, Inc. These values follow the concept of "High End Exposure" (HEE) exposure assumptions (US EPA, 1992, Exposure). Second, we calculate the RBCG as a distribution of BaPeq based on distributions for exposure assumptions taken from the refereed literature and/or as developed for this project (US EPA, 1992, Exposure). Third, we calculate a single RBCG based on deterministic default "Reasonable Maximum Exposure" (default RME) exposure assumptions proposed by US EPA Region V (US EPA, 1989, HHEM; US EPA, 1991, Default).

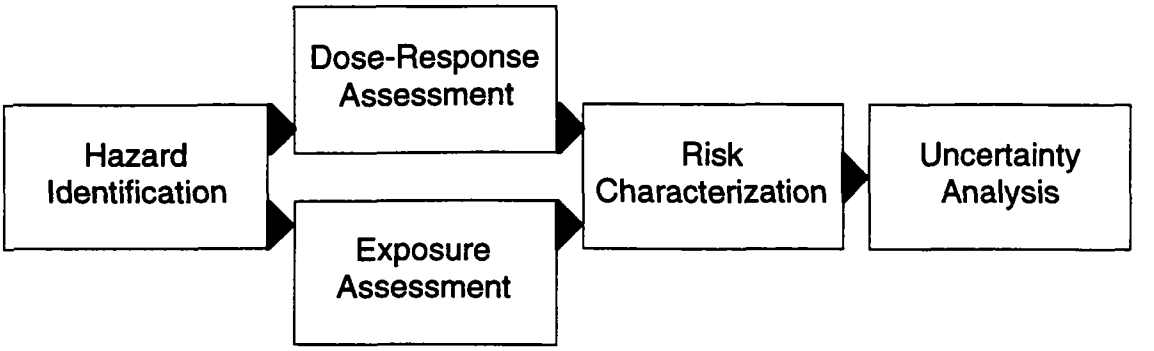
The three sets of calculations differ only in their exposure assumptions; all the calculations use (i) the same toxic potencies listed US EPA's Integrated Risk Information System (IRIS) database for benzo(a)pyrene (BaP) and (ii) the same (or more stringent) policy on target risk as adopted by US EPA's contractor, Ecology & Environment (E&E, 1995). While this report does not contain calculations using exposure assumptions based on the concept of Central Tendency (CT) exposure, we will provide them upon request.

The two sets of deterministic exposure assumptions in this report differ because Celotex and AlliedSignal rely primarily on the latest Agency "Guidelines for Exposure Assessment" published in the Federal Register (US EPA, 1992, Exposure) to develop

RME assumptions as point values using information appropriate for the neighborhood in Chicago. In contrast, Region V relies primarily on earlier Agency guidance in interim final directives (especially, US EPA, 1991, Default; and US EPA, 1989, HHEM) to develop default RME assumptions as point values using "bounding estimates." The terms "high end" and "bounding estimate" are defined in the Federal Register (US EPA, 1991, Exposure) and draw on recommendations and findings by the Agency's Science Advisory Board (US EPA, 1992, SAB). In short, Celotex and AlliedSignal develop assumptions that -- in combination -- are "a plausible estimate of the individual risk for those persons at the upper end of the risk distribution" (US EPA, 1992, Exposure, p. 22921), while Region V develops assumptions that -- in combination -- create a "bounding estimate" of risk.

Of course, none of the exposure variables in this report are truly point values; they are really distributions expressing the range and probability of occurrence. Using full-information methods, the probabilistic exposure assessment in this report shows how to propagate the range and probability in the exposure variables using a powerful and general mathematical tool called Monte Carlo simulation (Morgan & Henrion, 1990; Fishman, 1996). First used by mathematicians and physicists 50 years ago, Monte Carlo simulations are now used routinely in operations research, weapons design for national defense, and troop deployment -- not to mention mathematics, physics, chemistry, biology, oceanography, physiology, traffic engineering, bridge design, airport operation, investment banking, insurance, and other disciplines -- to do calculations for which deterministic methods give incorrect, partial, or misleading answers (Rubinstein, 1981; Morgan, 1984). The National Academy of Science (NAS), the National Council on Radiation Protection (NCRP), the American Industrial Health Council (AIHC), US EPA's Headquarters, and the Agency's Science Advisory Board have all endorsed Monte Carlo methods as a valid and powerful way to propagate distributions through dose equations and as a way to avoid excessively compounding conservatisms in health risk assessments (NAS, 1994; NCRP, 1996; AIHC, 1994; US EPA, 1992, Exposure; US EPA, 1992, SAB). Many, if not all, of the Agency's Regional Offices have now accepted one or more risk assessments for civilian or military projects using the Monte Carlo methods used in this report (e.g., US EPA ,1994, MC; US EPA, 1995, TG).

Following the standard practice in risk assessment as established by the National Academy of Sciences, the US EPA, and various state agencies, Alceon completed the core of this human health risk assessment in accordance with this simplified diagram:




In the first step, Hazard Identification (Section 4), Alceon identifies those chemicals present sufficient quantities or concentrations to pose a risk. The chemicals chosen for further study are called the Study Chemicals. The next two steps happen in parallel. In the Dose-Response Assessment (Section 5), Alceon reviews toxicological information about each Study Chemical to estimate adverse health effects associated with different doses. In the Exposure Assessment (Section 6), Alceon uses mathematical models along with assumptions or probability distributions to estimate the frequency, intensity, and duration of exposures that different groups of people may experience while living, working, playing, or visiting the Study Area. In this step, we consider potential exposures through various routes. In the fourth step, Risk Characterization and Quantitative Uncertainty Analysis (Section 7), we estimate the probability and/or magnitude of adverse health effects, if any, from the exposures of different groups of people to the Study Chemicals. In the last step, Uncertainty Analyses (Section 8), we discuss the many sources of variability and/or uncertainty in the methods, models, and assumptions, and we discuss our approach to understanding or quantifying the magnitude of the conservative assumptions in the overall approach.

The text concludes with a summary of the risk analyses (Section 9), a statement of limitations (Section 10), a list of abbreviations and acronyms (Section 11), and a list of references (Section 12). A map and all figures and tables appear following the text. Appendices present supporting information.

Throughout this report, we present concentrations in soils in milligrams per kilogram (mg/kg, equivalent to parts per million or ppm).

Throughout this report we present references in the following form: name of author, year of publication, and, when we refer to more than one work by the same author in the same year, an abbreviation of the title which is unique to that reference. For example,

we refer to the US EPA's "Human Health Evaluation Manual" as "US EPA, 1990, HHEM."

David E. Burmaster, Ph.D., wrote this report with help from Andrew M. Wilson and Steven J. Luis. Louis Anthony "Tony" Cox, Jr., Ph.D., and Brian H. Magee, Ph.D., wrote several of the appendices. David Burmaster visited the industrial property and surrounding neighborhood during the week of 17 April 1995.

# 2. Overview of Probabilistic Methods

This section provides an overview of probabilistic methods and their application to health risk assessment.

## 2.1 Why Probabilistic Methods are Useful

Traditional (deterministic) risk assessment methods select single point estimates (including many values ≥90th or ≥95th percentile of the range) for each exposure variable, such as soil ingestion rate, exposure frequency, absorption, body weight, etc. A major drawback of this approach is that it does not include either (i) the variability found in nature, physiology, and behavior or (ii) the uncertainty inherent in our knowledge. Combining a set of point values -- some typical, some conservative, and some very conservative -- yields a point estimate of risk that falls at an unknown percentile of the full distribution of risk. The degree of health-protectiveness afforded by the point estimate of risk, while believed to be highly conservative for reasons explained later in this report, is unknown and cannot be quantified by any deterministic method (NCRP, 1996).

Rather than select one point estimate from a range of values, a better approach is to use the entire range of possible values in the risk calculation (NCRP, 1996; Burmaster & Anderson, 1994). This can be accomplished using probabilistic techniques, such as Monte Carlo analysis, which uses distributions for input variables and generates distributions of outputs (Rubinstein, 1981; Morgan, 1984; Morgan & Henrion, 1990; Burmaster & von Stackelberg, 1991). The resulting distribution provides a full characterization of risks or cleanup goals and corresponding percentiles, which is considerably more useful and informative than a single value. The use of a probabilistic approach, rather than the traditional point estimate approach, is now recognized within the risk assessment community as more accurate and realistic (US EPA, 1992, RC; US EPA, 1992, Exposure; US EPA, 1994, MC; US EPA, 1994, MC; US EPA, 1995, TG; US EPA, 1995, GRC; Browner, 1995; US EPA, 1995, EFH2; MA DEP, 1993; Morgan & Henrion, 1990; NCRP, 1996).

## 2.2 Strengths of Probabilistic Methods

There are several reasons why probabilistic methods, such as Monte Carlo analysis, are appropriate for assessing health risks (Thompson et al, 1992; Morgan & Henrion, 1990;

US EPA, 1992, RC; US EPA, 1995, GRC; US EPA ,1994, MC; US EPA, 1995, TG; Harris & Burmaster, 1992; Burmaster & Harris, 1993).

First, the standard definition of risk (see, e.g., Webster's Dictionary) starts: "risk is the chance [meaning, probability] of harm...." At its root, risk assessment is the assessment of the probability of an adverse outcome. However, in the past, deterministic risk assessors routinely eliminated all aspects of probability in the deterministic methods published in its standard Superfund guidance manuals (US EPA, 1989, 1991). Probabilistic methods reintroduce the most basic notion of risk assessment -- probability -- into the practice of risk assessment.

Second, probabilistic methods quantify two very basic and important concepts -- variability (V) and uncertainty (U) -- at the same time as they estimate risks. In this report, we adopt the most common definitions for the terms. Variability represents heterogeneity or diversity in a well-characterized population, usually not reducible through further measurement or study. Uncertainty represents ignorance -- or lack of perfect information -- about a poorly-characterized phenomenon or model, sometimes reducible through further measurement or study. With probabilistic methods, risk assessors can quantify and understand the importance of V and U in a study.

Third, probabilistic methods are "full information" methods. In a deterministic risk assessment, the risk assessor destroys information about a site and the behavior of people to select a single number purported to represent the entire phenomenon. For example, in a deterministic risk assessment, the analyst discards most of the information known about children's body weights and instead uses a single value, say, 20 kg to represent the range of body weights in a particular age group. In a probabilistic risk assessment, the analyst does not destroy any information about a phenomenon. In a probabilistic risk assessment, the analyst does not destroy important correlations and dependencies among the variables. Thus, a probabilistic risk assessor uses the full range of values in an appropriate statistical framework.

Fourth, the need to include variability and uncertainty in a risk assessment is inherently understandable and acceptable to members of the public. The public understands that all people do not weigh the same or drink the same amount of water every day. Furthermore, it makes common sense to the lay public to incorporate this variability into an analysis, rather than assume that all humans have identical physiology and behaviors. While lay people do not understand all the mathematics inherent in

probabilistic methods, they quickly embrace the ideas inherent in variability and uncertainty.

## 2.3 Mechanics of Probabilistic Risk Assessment

In a traditional or deterministic risk assessment, the risk assessor builds a spreadsheet to estimate the risk through numerical computation. To simplify the discussion, consider a formula of the general form of Eqn 1:

$$\text{Risk} = f(X, Y, Z) \qquad \text{Eqn 1}$$

where Risk is a function of only three input variables, X, Y, and Z. In a deterministic risk assessment, the analyst picks a single number for X, a single number for Y, and a single number for Z. After the risk assessor inputs the function f( X, Y, Z ), the spreadsheet calculates the single value for Risk in Eqn 1.

In a probabilistic risk assessment, the basic risk equation remains the same:

$$\underline{\text{Risk}} = f(\underline{X}, \underline{Y}, \underline{Z}) \qquad \text{Eqn 2}$$

but, each of the input variables in Eqn 2 now is now underscored to indicate that each is a random variable, i.e., a variable that can take different values within a range of values as described by a probability distribution. In a probabilistic risk assessment, the analyst takes not just a single number for each input but instead develops a whole distribution of values that represent the variability and/or the uncertainty in the input. After the risk assessor inputs the function f( X, Y, Z ), the spreadsheet -- along with auxiliary software -- computes a full distribution of Risk as follows. In a first iteration, the computer picks a random value from each of the three input distributions and computes a single value of Risk. In a second iteration, the computer picks a different random value from each of the three input distributions and computes a different single value of Risk. As the computer performs additional iterations, the software stores each of the intermediate results and displays them as a histogram of Risk. After many iterations (often 10,000 or more iterations), the histogram of Risk values converges to the distribution of Risk sought by the analyst.

In this analysis, we used commercially available software for the calculations: Microsoft's Excel V4.0 on the Macintosh as the spreadsheet and Decisioneering's Crystal Ball V2.0 as the Monte Carlo simulation program (Decisioneering, 1992). We

used 20,000 iterations for all loops in the final calculations, a number large enough to demonstrate solid numerical stability.

## 2.4 Commonly Used Probability Distributions

This section provides an overview of the continuous distributions used in the probabilistic sections of this risk assessment. These and other distributions are discussed in many text books; the presentation and discussion in "Statistical Distributions, Second Edition" (Evans, Hastings, & Peacock, 1993) is particularly helpful.

### 2.4.1 Background Information

In full-information risk assessments, analysts commonly use continuous distributions (as opposed to discrete distributions) to represent exposure variables. In mathematics and statistics, there are many different but equivalent ways to express the same probability distribution in graphs or algebra, including but not limited to (i) probability density functions (PDFs), (ii) cumulative distribution functions (CDFs), and (iii) lists of percentiles (a different way to specify a CDF). Information presented in one form may be converted into another form; for example, a CDF is the integral of a PDF. To make this report as accessible as possible, we often use algebraic and graphical PDFs, but the same information can be equally well presented using CDFs and tables of various percentiles.

Statisticians often categorize probability distributions into families based on shape or mathematical properties. For example, normal distributions always follow the familiar bell-shaped curve when represented as PDFs or a less-familiar symmetric sigmoid curves when presented as CDFs. Some families of distributions, e.g., normal distributions, have been studied extensively because they occur throughout science and engineering. These common distributions are called parametric distributions because they can be completely specified by one, two, or a small number of parameters. For example, a normal distribution can be completely specified by two parameters, usually chosen as the mean and the standard deviation. Other parametric distributions include the uniform, triangular, lognormal, beta, and exponential distributions (Evans, Hastings & Peacock, 1993; Mood et al, 1974; Benjamin & Cornell, 1970; Parzen, 1960).

Nonparametric distributions also arise in practice for any one of several possible reasons. First, some situations require the development of a nonparametric distribution, i.e., a probability distribution not found in one of the common families (Green &

Silverman, 1994; Tarter & Lock, 1993). Such a distribution is also called a custom or empirical distribution. Second, sums and products of parametric distributions are rarely parametric distributions. For example, we compute products and sums of distributions of many random variables in this report. While these products and sums are valid probability distributions, they do not come from a parametric family, so we call them nonparametric, empirical, or custom distributions.

Every distribution, whether parametric or nonparametric, can be described by certain summary statistics, including the minimum and the maximum value. Most people are familiar with the two most common summary statistics -- the arithmetic mean (also called the average or expected value) and the standard deviation. In this report, we also use the median or 50th percentile, the value that occurs midway in a distribution, with half the values falling below the median and half falling above the median. In other parts of the report, we also use other percentiles, e.g., the 20th percentile, the 75th percentile, and/or the 95th percentile of a distribution.

### 2.4.2 Parametric Distributions Used in this Report

Uniform Distribution

Also called the rectangular distribution, the uniform distribution is used to represent a random variable that is equally likely to take any value between a minimum and a maximum (Evans, Hastings & Peacock, 1993). The uniform distribution is illustrated in Figure 1. A uniform distribution (with two parameters representing the minimum and the maximum) can be written as:

$$\text{Uniform}(\text{min}, \text{max}) = \text{Uniform}(2, 6) \qquad \text{Eqn 3}$$

This random variable is equally likely to take any value between the minimum of 2 and the maximum of 6. In this example, the random variable has a mean value of 4.

Risk assessors commonly use a uniform distribution to represent a random variable chosen on the basis of professional judgment when only the minimum and maximum are known. The distribution may represent variability, uncertainty, or a combination of both.

Triangular Distribution

The triangular distribution is also used to represent a random variable that takes values between a minimum value and a maximum value. However, with the triangular distribution, the values taken are not equally probable within the range (Evans, Hastings & Peacock, 1993). Instead, some central values are more likely than extreme values, with the overall relative frequency governed by a peaked distribution formed by two straight lines. Figure 1 illustrates the triangular distribution. A triangular distribution (with three parameters) can be written as:

Triangular(min, mode, max) = Triangular(2, 3, 7) Eqn 4

This random variable has a minimum of 2, a mode of 3, and a maximum of 7. The mode, also called the peak, is the single most likely value for the random variable. In this example, the random variable also has a mean of 4.

Risk assessors commonly use a triangular distribution to represent a random variable chosen on the basis of professional judgment when the probability has a single mode between a known minimum and maximum. The distribution may represent variability, uncertainty, or a combination of both.

Normal Distribution

The normal distribution is used to represent a random variable without a fixed minimum or a fixed maximum. Typically, normal distributions arise from "additive" processes (Evans, Hastings & Peacock, 1993). Normal distributions also arise naturally from repeatedly computing the averages of independent samples drawn from any type of distribution, not just a normal distribution. Figure 1 illustrates the PDF and the cumulative distribution function (CDF) for the normal distribution. A normal distribution (with two parameters) can be written as:

Normal(mu, sigma) = Normal(2, 1) Eqn 5

In this example, the normal random variable has a mean of 2 and a standard deviation of 1, meaning that approximately 68 percent of the area under the curve of the PDF occurs within the interval from (mu - one standard deviation) to (mu + one standard deviation). All normal distributions are symmetrical in shape -- it is equally likely for a value to fall above the mean of the distribution as below the mean.

LogNormal Distribution

The lognormal distribution is used to represent a random variable with a fixed minimum of zero but with no fixed maximum. Thus, a lognormal random variable describes a positive random variable. Typically, lognormal distributions arise from "multiplicative" or "dilution" processes such as the dispersion of a contaminant in water or soil. Lognormal distributions occur frequently throughout physics, chemistry, biology, and toxicology. Figure 1 illustrates the PDF and the CDF for the lognormal distribution. A lognormal distribution (with two parameters) can be written as (Evans, Hastings & Peacock, 1993; Aitchison & Brown, 1957; Crow & Shimizu, 1988):

LogNormal(mu, sigma) = exp[Normal(mu, sigma)] Eqn 6

= exp[Normal(2, 1)]

In this example, the natural logarithm of the random variable has a mean of 2 and a standard deviation of 1. All lognormal distributions are positive and asymmetric -- they never have negative values and they always have a (long) tail to the right. Risk assessors use lognormal distributions to describe many different types of random variables, including exposure point concentrations (Ott, 1995; Ott, 1990), body weights, skin areas, dietary intakes, breathing rates, and many other variables (US EPA, 1995, EFH2; AIHC, 1994; Anderson et al, 1984).

Exponential Distribution

The exponential distribution is commonly used to represent a random variable for the "time to failure" for a particular phenomenon, always a positive quantity. For example, exponential distributions are excellent models for the length of time that a new fluorescent tube will light before failing. Exponential distributions always have one parameter (often called lambda) that specifies the shape and location of the distribution and a long tail to the right (Evans, Hastings & Peacock, 1993). Figure 1 shows an exponential distribution:

Exponential(lambda) = Exponential(10) Eqn 7

For an exponential distribution, the mean = 1 / lambda. Recently, investigators have found that exponential distributions are excellent models for the length of time that a

person lives in a particular house and the length of time that a person keeps a certain job (Israeli & Nelson, 1992; Shaw & Burmaster, 1995).

Beta Distribution

The two-parameter beta distribution is used to represent a positive random variable with a fixed minimum of zero and a fixed maximum of one. Beta distributions are often used in risk assessments. For example, the fraction of a chemical absorbed in the human gut ranges between zero and one. Figure 1 shows the PDF for a particular beta distribution. A beta distribution has two parameters (often called a and b) that specify the shape and location of the distribution. A beta distribution can be written as:

$$\text{Beta}(a, b) = \text{Beta}(2, 4) \qquad \text{Eqn 8}$$

In this example, the mean = a / (a + b) = 1/3. If the two parameters are equal (a = b), then the beta distribution is symmetric. Otherwise, the distribution may be skewed to the left or the right depending on the values of the two parameters.

With two additional parameters, called c and d, the analyst can scale and translate a two-parameter beta distribution into a four-parameter beta distribution as follows:

$$\text{Beta4}(a, b, c, d) = \text{Beta}(a, b) \bullet c + d \qquad \text{Eqn 9}$$

This new random variable has a minimum value of d and a maximum value of (c+d).

# 3. Description and Background of the Site

The Streamlined Human Health Risk Evaluation (Ecology & Environment, 1995, Report) provides a description and history of the Celotex area.

## 3.1 Description of the Site

The industrial property is located at 2800 South Sacramento Avenue in Chicago, Illinois. The map (from E&E, 1995, Report) shows the location of the property. The Site, including the 18-acre industrial property and a 6-acre property to the south (which is currently used for storage of trucks), is located in a mixed industrial, commercial, and residential area. It is bounded on the east by the Cook County Correctional Facility, on the south by the Chicago Fire Department, Bureau of Support Services, and Farley Candy Company, and on the north and west by residences and by the Atchison, Topeka, and Santa Fe Railroad Lines.

Industrial buildings on the industrial property were demolished and removed in 1993. Some concrete slabs and remnants of foundations remain. Following demolition, a soil cover was placed on the property. The property is surrounded by a chain-link fence.

## 3.2 History of the Property

According to file information, the area was a farmstead before 1912. Little is known about the use of the area before this time.

Asphalt roofing products were manufactured at the property from 1912 until 1982. Information provided by Celotex and AlliedSignal indicates that the Barrett Company and Allied Chemical Corporation operated a coal tar distillation and roofing plant at the facility from 1912 until 1970. The Barrett Division of Allied Chemical Corporation was purchased by Jim Walter Corporation in 1967 and a portion of the property was transferred to the Celotex Corporation (Celotex), a subsidiary of the Jim Walter Corporation. Allied Chemical Corporation sold four additional parcels of the facility's property to Celotex between 1972 and 1975. In 1975, Allied Chemical Corporation sold the tar plant property to Service Welding and Cleaning Company (Service Welding) and subsequently leased a building from Service Welding to operate a sealer plant. Allied Chemical Corporation closed the sealer plant in 1977 and in 1980 Celotex bought all property that was purchased by Service Welding.

The Illinois Environmental Protection Agency (IEPA) completed a PA of the property in 1989. Subsequent field investigations included an SSI in 1991, an ESI in 1992, and an LSI in 1993. In addition to these studies, an US EPA Technical Assistance Team assessed the area in 1993.

On 8 November 1994, US EPA issued a Special Notice of Liability and 104(e) Information Request Letter to Celotex and AlliedSignal. The special notice letter requested that the PRPs sample to determine the extent and degree of PAH contamination in the vicinity of the industrial property, prepare an EE/CA to address contamination in residential areas that exceeded cleanup goals, and initiate investigation of the area. The PRPs submitted a Support Sampling Plan (SSP) (ERM, 1995) to US EPA to investigate the extent of site-related contamination within 2,500 feet of the fenced property.

## 3.3 Previous Calculations of Risk-Based Cleanup Goals

### 3.3.1 Ecology & Environment, October 1995

In a report dated October, 1995, Ecology & Environment (E&E) estimated point values for Risk-Based Cleanup Goals (RBCGs) for benzo(a)pyrene equivalents (BaPeq) in surface soils at residences near the industrial property (E&E, 1995, Report). We summarize their methods and findings here.

E&E selected PAHs based on "evaluation of analytical information provided by IEPA, and consideration of potential off-site migration pathways of contaminants as a result of prior site operations." Following common practice, E&E segregated the PAHs into two groups: carcinogenic and noncarcinogenic PAHs. To simplify calculations, the carcinogenic PAHs were converted to benzo(a)pyrene (BaP) equivalents (using the relative potencies reported in US EPA, 1993, PAHs). E&E selected pyrene to represent the noncarcinogenic PAHs due to its high noncarcinogenic toxicity.

Based on a preliminary characterization of the exposure setting and local population, E&E identified one exposure scenario, that of residents living in the vicinity of the industrial property. Three pathways were identified: incidental soil ingestion, inhalation of fugitive dust, and dermal contact with soil. E&E made numerous conservative assumptions to characterize each of these pathways for two exposure cases: the default reasonable maximum exposure (default RME) and central tendency (CT) exposure.

E&E did not estimate cancer risks due to dermal exposure because US EPA has not published toxicity values for dermal exposure. Therefore, only ingestion and inhalation were evaluated for cancer risk.

In accordance with the National Contingency Plan (US EPA, 1990, NCP), E&E selected $10^{-4}$ as the target Incremental Lifetime Cancer Risk (ILCR) and 1 as the target Hazard Index (HI) for carcinogenic and noncarcinogenic PAHs, respectively. E&E then estimated point values for the RBCGs for carcinogenic and noncarcinogenic PAHs in soils at residential properties.

There are at least two important conclusions to be drawn from the results of E&E's analysis. First, E&E showed that the carcinogenic PAHs are approximately two orders of magnitude (a factor of ~100) more important than are the noncarcinogenic PAHs in estimating cleanup targets. We agree with E&E's result on this point. From E&E's result, Alceon concludes that it is only necessary to consider carcinogenic PAHs when deciding the proper risk management strategy for the residences near the industrial property.

Second, E&E argued that the dominant pathway for exposure is the pathway for incidental ingestion of soil. According to the formulas and input values chosen by E&E, the second pathway -- the inhalation of fugitive dust -- is less important than the soil ingestion pathway for both carcinogenic and noncarcinogenic PAHs by more than four orders of magnitude (a factor greater than 10,000). In addition, citing guidance from the US EPA, E&E did not quantify any risks from exposures via dermal exposure.

With these methods and assumptions, E&E estimated these deterministic point values for RBCGs for surface soils at residences near the industrial property: (i) 8.6 mg/kg BaPeq based on default RME exposure assumptions, and (ii) 72 mg/kg BaPeq based on CT exposure assumptions.

### 3.3.2 Alceon Corporation, January 1996

In a report dated January 1996, Alceon estimated deterministic point values for RBCGs for BaPeq in surface soils at residences near the industrial property (Alceon, 1996, DRA). We summarize our methods and findings here.

Based on E&E's earlier report, and based on our own calculations, Alceon excluded the fugitive dust pathway and the dermal contact pathway from further consideration and

focused exclusively on the soil ingestion pathway in the risk assessment. Following the precedent used in the E&E report, Alceon used $10^{-4}$ as the target for incremental lifetime cancer risk.

With these methods and assumptions, Alceon estimated 29.5 mg/kg BaPeq as the deterministic point value as the RBCG for surface soils at residences near the industrial property.

### 3.3.3 Alceon Corporation, February 1996

In a report dated February 1996, Alceon estimated a probability distribution for the RBCG for BaPeq in soils at residences near the industrial property (Alceon, 1996, PRA). We summarize our methods and findings here.

Based on E&E's earlier report, and based on Alceon's earlier deterministic report, Alceon completed a risk assessment using probability distributions for exposure variables and a fixed value for the toxic potency of BaP. Following the precedent from Ecology & Environment, and paralleling our earlier deterministic report, we included only exposures via inadvertent ingestion, and we excluded exposures via fugitive dust and dermal contact. Again, using the same target risk as US EPA's contractor, we defined an acceptable distribution of risk as having, simultaneously, a 95th percentile of risk equal to or less than $10^{-4}$ risk (this parallels US EPA's policy on RME risk), and a 50th percentile (or median) of risk equal to or less than $10^{-5}$ risk (this parallels US EPA's policy on CT risk). Together, these two simultaneous constraints create a risk management policy that is more stringent than the one used by Ecology & Environment.

With these methods and assumptions, Alceon calculated the RBCG for surface soils at residences near the industrial property as a (truncated) lognormal distribution, summarized as follows: First, no single measurement of BaPeq may exceed 100 mg/kg at any location. Second, the set of all Exposure Point Concentrations of BaPeq must simultaneously meet all of these constraints developed from the truncated lognormal distribution:

| | | | |
|---|---|---|---|
| minimum | = | zero | mg/kg BaPeq |
| 10th percentile | ≤ | 6.3 | mg/kg BaPeq |
| 20th percentile | ≤ | 9.4 | mg/kg BaPeq |

| | | | |
|---|---|---|---|
| 30th percentile | ≤ | 12.8 | mg/kg BaPeq |
| 40th percentile | ≤ | 16.1 | mg/kg BaPeq |
| 50th percentile | ≤ | 20.7 | mg/kg BaPeq |
| 60th percentile | ≤ | 26.1 | mg/kg BaPeq |
| 70th percentile | ≤ | 33.5 | mg/kg BaPeq |
| 80th percentile | ≤ | 43.5 | mg/kg BaPeq |
| 90th percentile | ≤ | 59.3 | mg/kg BaPeq |
| 95th percentile | ≤ | 73.7 | mg/kg BaPeq |
| maximum | ≤ | 99.9 | mg/kg BaPeq |

These constraints on the distribution of concentrations of BaP equivalents create a distribution of EPCs that meets the definition of an acceptable distribution of risk.

### 3.4 Purpose of this Report

Since February 1996, Celotex, AlliedSignal, and US EPA Region V have had several discussions about the three risk assessment reports just summarized. In these discussions, US EPA Region V has said that it no longer considers it appropriate for the risk assessment to exclude exposures via dermal contact. Thus, the first purpose of this new report is to include and quantify exposures via the dermal pathway.

In addition, through the discussions, Celotex, AlliedSignal, and Region V have agreed on point values for some exposure variables but continue to differ on other exposure variables. Thus, the second purpose of this report is to develop deterministic RBCGs for BaPeq based on the deterministic assumptions now proposed by Celotex and AlliedSignal and those now proposed by US EPA Region V.

The third purpose of this report is to develop the RBCG for BaPeq as a distribution using full information methods, i.e., probabilistic methods. Since none of the exposure variables are truly point values, the probabilistic calculations put the risk assessment on a sound theoretical basis by treating the key exposure variables as the distributions that they truly are. Monte Carlo simulations propagate the variability in the input variables in

a mathematically correct way, and the simulations include the numerically important correlations (Smith et al, 1992) but avoid the excessive compounding of conservatisms inherent in default RME methods (Harris & Burmaster, 1992; Burmaster & Harris, 1993).

Finally, for convenience, appendices to this document compile all the background materials submitted by Celotex and AlliedSignal and its contractors under one cover.

# 4. Hazard Identification

## 4.1 Definition of the Study Area

We define the Study Area to be the residential neighborhood(s) in the vicinity of the industrial property. Residences in this area are primarily located to the north and west of the property. As in E&E's report, we limit the population of potentially exposed individuals to those residents living within the 2,500-ft sampling radius defined by ERM (1995, DSR). Since the industrial property itself is not residential, the exposure scenarios considered here are not applicable and an additional assessment would be required to determine appropriate RBCGs for the industrial property itself.

In Appendix A, we present the measurements of the concentrations of carcinogenic polycyclic aromatic hydrocarbons (cPAHs) -- as expressed in terms of benzo(a)pyrene equivalents (BaPeq, in mg/kg, equivalent to ppm) -- for 49 soil samples representing "urban background concentrations" for this project. As explained further below, the 49 measurements range from 0.7 mg/kg BaPeq for a sample in Douglas Park to 26.0 mg/kg BaPeq for a sample some 1,500 to 2,500 ft north of the industrial property. Taken together, these 49 samples provide a statistical population of measurements against which other populations of measurements may be compared using nonparametric tests such as the Wilcoxon Rank Sum test or the Kolmogorov-Smirnov test.

Using powerful statistical methods, Louis Anthony Cox, Jr. has demonstrated that the spatial pattern of the concentrations in these 49 samples are unrelated -- with 95 percent confidence -- to the industrial property (see Appendix B).

As shown by the lognormal probability plot in Appendix A, these 49 samples are well described by this lognormal distribution (Gilbert, 1987; Draper & Smith, 1981):

$$\ln[\text{BaPeq}] = \text{Normal}(\mu, \sigma)$$

which is equivalent to

$$[\text{BaPeq}] = \exp[\ \text{Normal}(\mu, \sigma)\ ]$$

$$= \exp[\ \text{Normal}(\ 1.01, 0.61)\ ]\ \text{in mg/kg}$$

This lognormal distribution for BaPeq in background soils has an arithmetic mean equal to ~3.31 mg/kg and a 95$^{th}$ percentile equal to ~7.49 mg/kg.

It is essential to treat this lognormal distribution as a full mathematical object because it is the object of interest. While the distribution presents the full information available in the background soil samples, any single value drawn from the distribution -- or any single statistic summarizing the distribution -- necessarily destroys information present in the 49 laboratory measurements. Thus, when comparing the soil concentrations in a "treatment" area to the soil concentrations in a "background" area, it is essential to compare the full distribution of measurements for the "treatment" area to the full distribution of measurements for the "background" area before drawing any inferences about the similarities or differences between the two areas (e.g., Mood et al., 1974).

To put these numbers in context, we present information from two authoritative studies of urban background concentrations of PAHs in surface soils.

- Bradley et al. (1994) report the results from the measurement of 60 surficial, nonindustrial soil samples from three New England cities. These 60 samples had a range from 0.26 mg/kg BaPeq to 21.31 mg/kg BaPeq, with an arithmetic mean concentration of 2.44 mg/kg BaPeq and a 95-percent UCL on the arithmetic mean of 3.32 mg/kg BaPeq.

- The US Agency for Toxic Substances and Disease Registry (ATSDR, 1990) reports the results from hundreds of measurements of background concentrations of PAHs in urban soils. In particular, applying US EPA's provisional guidance to ATSDR's values reported in Table 5-5 (page 148), we find that concentrations of BaPeq range from < 1 mg/kg to ~19 mg/kg in urban soils.

These two authoritative sources on urban background soil concentrations demonstrate that range and distribution of the site-specific concentrations of BaPeq in the 49 background surface soil samples analyzed in Appendix A fit well within the typical range and distribution of BaPeq measurements in cities across the country.

## 4.2 Selection of the Study Chemicals

As discussed earlier, we select cPAHs as the Study Chemicals (Table 1) since they dominate the calculations of RBCGs (E&E, 1995, Report).

## 4.3 Physical-Chemical Properties of the Study Chemicals

Table 2 summarizes the key physical and chemical properties of the Study Chemicals. We tabulated these values from databases compiled by the US EPA (US EPA, 1988, and US EPA, 1994, HSDB). Taken together, the physical-chemical properties in this table describe the tendencies of a particular compound to move and/or accumulate in various environmental media (Mackay et al., 1992; Verschueren, 1983).

A review of the physical and chemical properties of the PAHs in the Study Area reveals that the PAHs have: (i) moderate to high molecular weights, (ii) low water solubilities, (iii) low vapor pressures, (iv) low Henry's Law constants, (v) moderate to high water-carbon partition coefficients, and (vi) moderate to high octanol-water partition coefficients. These properties show that the PAHs included as Study Chemicals for this risk assessment generally adsorb strongly to soils, migrate slowly if at all in ground water, and do not readily volatilize into soil gas or the atmosphere.

## 5. Dose-Response Assessment

Table 3 summarizes the toxicological properties of the Study Chemicals for use in estimating human cancer risks. A compound may have values for carcinogenic responses from exposures via ingestion and/or inhalation. All toxicity values in this report are based on lifetime exposures.

The toxicity profile for benzo(a)pyrene supporting Table 3 was downloaded in November 1995 from the US EPA's Integrated Risk Information System (IRIS) database (US EPA, 1995, IRIS). See the full profile in Appendix C.

The toxicity values in Table 3 are the most current and authoritative available from the US EPA. Specifically:

- if available, the table presents values from US EPA's IRIS database (US EPA, 1995).
- if the IRIS database does not include a particular value, the table presents a value from the US EPA's most recently published Health Effects Assessment Summary Tables (US EPA, 1994, HEAST), if available.
- in general, if toxicity values are not available from any of these sources, blanks appear in the table.

The US EPA frequently revises the toxicity values it publishes in the IRIS database and the quarterly HEAST. The toxicity values in this risk assessment are current as of the dates shown on the IRIS profiles in Appendix C, and as of the version of HEAST cited in the references.

Published toxicity values may be based on either an exposure dose or an absorbed dose. In estimating health effects, the dose and toxicity value must be concordant -- that is, a risk estimate should use either an exposure dose and a toxicity value based on exposure dose or an absorbed dose and a toxicity value based on absorbed dose. According to US EPA (US EPA, 1989, HHEM), most of the US EPA's published Reference Doses (RfDs) and some Cancer Slope Factors (CSFs) are based on exposure dose.

## 5.1 Carcinogenic Toxicity of the Study Chemicals

For each of the carcinogenic PAHs reported in the Study Area, Table 3 presents both ingestion and inhalation Cancer Slope Factors (CSFs) measured in units of risk per unit dose: inverse milligram of chemical per kilogram of body weight per day or $(mg/(kg \cdot day))^{-1}$. The larger the CSF, the more potent the compound. The US EPA estimates the CSFs based on the assumption that a nonzero dose causes a nonzero probability of carcinogenic response; that is, the CSF values are generally based on a linear nonthreshold dose-response model. The US EPA has not developed CSFs for dermal exposures.

When the US EPA assesses a compound for human carcinogenicity, the Agency gives a Weight-of-Evidence rating that reflects its confidence in the evidence of carcinogenicity. (See Table 4; US EPA, 1986, FR). The US EPA's Weight-of-Evidence categories generally parallel those developed by the International Agency for Research on Cancer (IARC). As described in the documentation supporting the IRIS database (US EPA, 1994-5), chemicals that give rise to cancer and/or gene mutations are generally classified by US EPA as follows: (i) Group A: Human Carcinogen; (ii) Group B1: Probable Human Carcinogen; limited human data; (iii) Group B2: Probable Human Carcinogen; sufficient evidence in animals and limited or no evidence in humans; and (iv) Group C: Possible Human Carcinogen. If insufficient tests for carcinogenesis or mutagenesis are available, the US EPA generally places the chemical in Group D: Not Classifiable as to Human Carcinogenicity. A fifth category, Group E: Evidence of NonCarcinogenicity in Humans, is rarely used. The US EPA usually publishes CSFs for chemicals classified as Group A, B1, B2, or C. Finally, the US EPA has recently proposed changes in the regulation of carcinogens, and the new "Cancer Guidelines" are still open to pubic comment (US EPA, 1996, CG).

Unless human data from occupational exposures are available, the published CSF is derived using a statistical model applied to the results of animal experiments, and, unless otherwise indicated, the published CSF for humans is extrapolated from the 95-percent upper confidence limit on the linear term of the linearized multistage model fit to the animal data. By assumption and methodology, CSFs estimated this way are uncertain values, and most experts believe that they are generally conservative (i.e., they tend to overestimate risk).

Relative potencies for the carcinogenic PAHs are listed in Table 5. Following standard risk assessment practice to streamline calculations, we consider health risks due to BaP equivalent concentrations. Benzo(a)pyrene equivalents for a contaminated medium are calculated as a weighted sum by multiplying concentrations of nonbenzo(a)pyrene PAHs by the relative potencies listed in Table 5. The resulting weighted concentrations are then summed together with the benzo(a)pyrene concentration to yield a benzo(a)pyrene equivalent concentration for the medium (US EPA, 1993, PAHs; see also US EPA, 1986, Mixtures).

The US EPA has not published any CSFs for exposures to any compounds via dermal pathways. However, based on conversations with Region V staff members, Celotex and AlliedSignal have agreed to include possible risks from dermal exposure in the calculations in this report by cross-assigning the CSFs for the ingestion pathway to the dermal pathway.

# 6. Exposure Assessment

This section describes the exposure scenarios considered in this risk assessment (US EPA, 1990, EFH).

## 6.1 Summary of Exposure Scenarios

### 6.1.1 Classification of Soil

IEPA and ERM (1995, DSR) reported surface soil data for samples collected in the top few inches of the soil. We assume that the wind has transported some PAHs from the industrial property and has deposited some of them on surface soils. We also assume that some of the PAHs now found in soils in the residential neighborhood come from other industrial, commercial, and/or residential sources (ATSDR, 1990).

PAHs in soils are bound to the soil matrix by chemi-adsorption; this physical-chemical bond increases in strength over time (GRI, 1995). For these reasons, exposure scenarios have been limited to surface soils (E&E, 1995, Report). Following this precedent, we consider only exposures to surface soils in the vicinity of the industrial property.

### 6.1.2 Populations and Exposure Pathways of Concern

This human health risk assessment considers exposures to people living in houses near the industrial property under current and future conditions (Table 6). We consider three age groups of people: children (ages 1 through 6 years), teenagers (ages 7 through 17 years), and adults (ages 18 years and older).

Each person could theoretically be exposed to Study Chemicals via any of the exposure pathways listed in Table 6:

- ingestion of soil outdoors (and dust indoors),
- inhalation of fugitive dust,
- dermal contact with soils,
- inhalation of soil vapors.

As noted above, PAHs exhibit little tendency to volatilize. For this reason, we believe inhalation exposures to soil vapors, and any risks resulting therefrom, to be negligible. We do not consider them further.

As discussed above, E&E (1995, Report) estimated incremental cancer risks associated with inhalation of fugitive dust for the population of residents. We have checked their calculations, and we agree that the estimate of incremental cancer risk due to inhalation of fugitive dust is more than four orders of magnitude (a factor of >10,000) less than the incremental cancer risk due to incidental soil ingestion. Therefore, we do not consider the inhalation pathway further.

As discussed in the E&E reported (October 1995), the US EPA has not published CSFs for any PAHs via dermal contact. Therefore, E&E did not evaluate cancer risks resulting from dermal exposure in their report, nor did Alceon evaluate possible cancer risks via dermal exposure in either of our earlier risk assessments (Alceon, 1996, DRA; 1996, PRA). However, based on conversations with Region V staff members, Celotex and AlliedSignal have agreed to include possible risks from dermal exposure in the calculations in this report by cross-assigning the CSFs for the ingestion pathway to the dermal pathway. It is assumed that exposure may occur via ingestion or dermal contact with PAHs in outdoor surface soils or indoor dust.

## 6.2 Exposure Variables

For each of the exposed age groups in the residential population, we present three sets of exposure variables. First, we present point values for the exposure variables for RME conditions as proposed by Celotex and AlliedSignal. Second, we present full-information distributions for the exposure variables as proposed by Celotex and AlliedSignal. Third, we present point values for the exposure variables for default RME conditions as proposed by Region V.

Table 7 shows the assumptions for each exposure variable. The first three columns of Table 7 list the name of the exposure variable, the algebraic symbol for the exposure variable, and its units. The next three columns list the point estimates proposed by Celotex and AlliedSignal for RME conditions, along with a citation to the source document. The next four columns of Table 7 list the distribution proposed by Celotex and AlliedSignal, including the probability density function (PDF) for the distribution, its support, its 95th percentile, and its median (50th percentile). The second page of the

table lists the point estimates proposed by Region V for default RME conditions in a similar format.

## 6.3 Detailed Exposure Scenarios

### 6.3.1 Exposure Frequency

#### 6.3.1.1 Total Exposure Frequency

Celotex, AlliedSignal, and Region V agree that residents living near the Celotex property are exposed for a total of 350 days per year. We use this point value in all the calculations, both deterministic and probabilistic. This value originates from the default values published among Interim Final Standard Exposure Factors (US EPA, 1991, Default) for the Superfund program nationwide. The value is based on the implicit assumption that each child, teen, and adult travels or vacations outside the neighborhood some 15 days each year.

#### 6.3.1.2 Exposure Frequency to Soils Outdoors

Celotex and AlliedSignal: Celotex and AlliedSignal consider that children, teens, and adults have 164, 177, and 167 days per year, respectively, of exposure to surface soils outdoors for RME conditions. These point values are chosen as the median values from the distributions developed on a site-specific basis as follows (see also Appendix D):

The frequency of exposure is a function of presence at a potential exposure point and activity while at the potential exposure point. In particular, since soil ingestion rates are based on the quantity of soil ingested per day of exposure, we estimate the number of days an individual is engaged in an activity that results in contact with surface soil in the neighborhood.

Activities at the residence vary widely. Typical indoor activities include sleeping, preparing and eating meals, watching TV, playing, doing housework. Typical activities outdoors include playing and gardening. Data describing the frequency with which individuals participate in these activities are scarce.

Lacking detailed information concerning activity patterns, we instead rely on a surrogate for activities that involve exposures to surface soils. This surrogate is based on the following reasoning: We know that weather strongly affects the nature and duration of

outdoor activities. Since weather data have been routinely recorded at both O'Hare and Midway Airports for many decades (US DoC, 1992), we use historical weather conditions at Midway Airport -- the closer of the two airports in Chicago -- as a surrogate for information concerning activity patterns outdoors.

Table 8 summarizes the average daily temperatures recorded at Midway Airport from 1961 to 1990 (US DoC, 1992). More specifically, the top two rows of data in Table 8 show the number of days per year that are at or above the stated temperature and -- by difference -- the number of days per year that are below the stated temperature. For example, in a typical year, there are 196 days with average temperatures ≥ 50 degF and 169 days with average temperature < 50 degF.

For the children, we assume that the average daily temperature strongly influences the fraction of days in a year on which a child plays outside and incidentally ingests some surface soil. In particular, we assume these breakpoints as shown in Table 8:

- For days when the average daily temperature is < 32 degF, we assume that no child incidentally ingests soils outdoors. On such cold days, the outdoor soils are frozen and/or covered with snow and ice.
- For days when the average daily temperature is < 40 degF, we assume that 5 percent of children incidentally ingest some surface soils outdoors.
- For days when the average daily temperature is < 50 degF, we assume that 20 percent of children incidentally ingest some surface soils outdoors.
- For days when the average daily temperature is < 60 degF, we assume that 70 percent of children incidentally ingest some surface soils outdoors.
- For days when the average daily temperature is ≥ 70 degF, we assume that 100 percent of children incidentally ingest some surface soils outdoors.

These assumptions, shown in Table 8, define a range (distribution) for the number of days per year that a child incidentally ingests surface soils or has dermal contact with surface soils outdoors. The median of this distribution is 164 days per year, the value chosen as the RME input for children in Table 7. Appendix D gives the full-information probability distribution for this exposure variable in the form of custom, piece-wise linear distribution.

For teens, we assume the breakpoints shown in Table 8. These assumptions reflect the fact that teens are more mobile and active outdoors than children, and so we assume higher frequencies of exposure for teens than for children. These assumptions for teens define a range for the number of days per year that a teen incidentally ingests surface soils or has dermal contact with soils outdoors. The median of this distribution is 177 days per year, the value chosen as the RME input for teens in Table 7. Appendix D gives the full-information probability distribution for this exposure variable in the form of custom, piece-wise linear distribution.

For adults, we assume the breakpoints shown in Table 8. These assumptions reflect the fact that adults are less mobile and active outdoors than teens, and so we assume frequencies of exposure for adults similar to the ones for children. These assumptions for adults define a range for the number of days per year that an adult incidentally ingests surface soils or has dermal contact with surface soils outdoors. We use the distribution for adults developed in Appendix D. The median of this distribution is 167 days per year, the value chosen as the RME input for adults in Table 7. Appendix D gives the full-information probability distribution for this exposure variable in the form of custom, piece-wise linear distribution.

Region V: For default RME conditions, Region V assumes that each child, teen, and adult has exposure to surface soils outdoors on 350 days each year, i.e., every day regardless of temperature or snow cover. Region V chooses 350 days per year as the default RME exposure frequency from the default values published among the Interim Final Standard Exposure Factors (US EPA, 1991, Default) for the Superfund program nationwide.

#### 6.3.1.3 Exposure Frequency to Dust Indoors

Celotex and AlliedSignal: For each age group, we estimate the number of days of exposure to dust inside the house as the difference between the total number of days of exposure each year (350 days per year) and the number of days of exposure to soils outside the house. For RME conditions, children, teens, and adults have 186, 173, and 183 days per year, respectively, of exposure to dust inside the home.

Region V: For each age group, Region V also assumes that the number of days of exposure to dust inside the house equals the difference between the total number of days of exposure each year (350 days per year) and the number of days of exposure to

soils outside the house. Since Region V assumes that each person has 350 days per year of exposure outside the house for default RME conditions, the Region assumes, in effect, that each person has zero days per year of exposure inside the house.

### 6.3.2 Exposure Duration

Table 7 shows the assumptions made by Celotex and AlliedSignal and by Region V about the duration of exposures for various age groups in the neighborhood.

Celotex and AlliedSignal: Celotex and AlliedSignal consider that children, teens, and adults have 6, 11, and 1 years of exposure, respectively, in the neighborhood. The value for total exposure duration for RME conditions (18 years) was chosen as the 90th percentile of the neighborhood-specific occupancy duration, as detailed in Appendix E.

The duration of exposure is limited to the period in which an individual lives in the neighborhood. When they were employees of the US Environmental Protection Agency, Israeli and Nelson (1992) estimated distributions of time of residence for different groups of US households based on data published by the Bureau of the Census. Israeli and Nelson report that the distribution for total residence time is essentially an exponential distribution with a different mean value for each housing group. An exponential distribution is completely characterized by that mean value (as a single parameter) and is highly skewed, with a long tail to the right.

Although Israeli and Nelson (1992) estimate distributions of residence time for households and not for individuals, they state that corresponding residence times for individuals are expected to be smaller. They also state that "[t]he values calculated here can be considered to represent *upper limits* of the expected time for individuals to live at the same residence." (emphasis added). Thus, our use of the distributions of residence time estimated by Israeli and Nelson (1992) results in conservative estimates of exposure duration for potentially exposed individuals.

US EPA Region V's contractor (Ecology & Environment, 1995, Letter) has stated "A review of 1990 housing population statistics for Chicago's South Lawndale Community Area (where the site is located) indicates that the census tract in which the site is located includes a significant amount of owner-occupied households (approximately 49 percent) (US Census Bureau, 1990)."

We accept 49 percent as the fraction (a point value) of owner-occupied houses in the neighborhoods surrounding the industrial property. From this, 51 percent is the fraction (a point value) of nonowner-occupied houses in the same neighborhoods (US BoC, 1991).

In Table IV, in a column titled "Average total residence time, T (years)", Israeli and Nelson (1992) show that the exponential distribution for "Owners" is characterized by a mean value of 11.36 years and that the exponential distribution for Renters is characterized by a mean value of 2.35 years. From these values, we estimate the RME total exposure duration as the 90th percentile of the neighborhood-specific occupancy as 18 years. (See also supporting materials in Appendix E.)

To estimate the exposure duration for each age group, we assumed that a potentially exposed individual spends his or her first years of life at the residence. LaGoy (1987) reports that children of age one year or less have little direct contact with soil. We note that the assumption of spending the first years of life at a residence near the property is health protective because the rate of soil ingestion is generally higher for children than for teens or adults. Also, other exposure factors (e.g., low body weight for children) combine to increase the dose received by children relative to older age groups.

For children, we assume that exposures to surface soil start at age 1 year. Therefore, we consider that RME exposure for children begins at age one year and continues through age 6 years, for a total of 6 years. This value is reported in Table 7.

For teens, the duration for RME exposure is 11 years (from age 7 years through age 17 years). We assume older children and teenagers are exposed for the time during which they reside in the vicinity of the industrial property. Subtracting exposure during childhood, this is a period of 11 years. Therefore, RME exposure for teens covers 11 years, as reported in Table 7.

For adults, we use subtraction to find that adults are exposed for 1 year for RME conditions.

Overall, we quantify RME exposures to a person for a full 18 years -- 6 years as a child, 11 years as a teen, and 1 year as an adult. This method is conservative because it assigns RME exposure to begin with the group (children) most likely to receive the highest dose. Appendix E gives the full-information probability distribution for this

exposure variable in the form of a compound distribution for the mixture of two exponential distributions. In each iteration of the Monte Carlo simulations, we conservatively assign the first years of exposure the child, then any remaining balance of exposure to the teen, and then any remaining balance of exposure to the adult.

Region V: For default RME conditions, Region V assumes that each person lives in the neighborhood for a total of 30 years, with this time allocation: 6 years for children, 11 years for teens, and 13 years for adults. Region V chooses 30 years as the default RME exposure duration from the default values published among the Interim Final Standard Exposure Factors (US EPA, 1991, Default) for the Superfund program nationwide.

### 6.3.3 Exposure via Ingestion of Outdoor Soils and Indoor Dust

Table 7 shows the assumptions made by Celotex and AlliedSignal and by Region V about the ingestion rate of outdoor soil and indoor dust for various age groups for the default RME conditions. Table 7 also shows the distributions used in the simulations.

#### 6.3.3.1 Ingestion of Outdoor Soil

Celotex and AlliedSignal: For RME conditions, Celotex and AlliedSignal assume that children, teens, and adults incidentally ingest 200, 100, and 100 mg/day of outdoor soils on a day of outdoor exposure. In the Monte Carlo simulations, Celotex and AlliedSignal assume the lognormal distributions listed in Table 7 for children, teens, and adults, as based on Thompson & Burmaster (1991).

For children, LaGoy (1987) and Hawley (1985) report that incidental soil ingestion can occur outdoors at any age, but is most prevalent among young children because young children (less than 3 or 4 years old) often mouth small objects.

Rates of incidental ingestion of soil outdoors by young children have been directly measured by several researchers. In particular, Calabrese et al. (1989) and Binder et al. (1986) have applied the Limiting Tracer Method (LTM) to the problem of estimating incidental soil ingestion rates in young children. In these studies, soil ingestion rates were estimated by analyzing feces for trace elements found in soils but not typically found in foods. Accounting for differences in experimental design, their results are consistent.

Based on the analysis by Binder et al. (1986), Thompson and Burmaster (1991) estimate distributions of incidental soil ingestion rates outdoors for young children. We use these results here. Thompson and Burmaster (1991) report that the data follow this lognormal distribution: exp[N( 4.13, 0.80 )] in mg/day. For RME conditions, the 90th percentile of this distribution is 173.1 mg/day. Rather than use the 90th percentile of the distribution in this risk assessment, we instead use 200 mg/day as the RME value for the ingestion rate for outdoor soil and indoor dust in accordance with US EPA guidance as the ingestion rate for children (see Table 7). This higher value is the 93rd percentile of the distribution of measured values.

For teens, LaGoy (1987) notes that soil ingestion rates for older children and teenagers have not been studied extensively. Both LaGoy (1987) and Hawley (1985) indicate, however, that soil ingestion rates are expected to decrease as children grow older because mouthing of objects decreases with age. LaGoy (1987) assumes incidental soil ingestion rates of children 6 to 11 years of age to decrease by at least 50 percent. LaGoy suggests that incidental ingestion rates are lower still for older children and teenagers. Based on LaGoy, we assumed the RME ingestion rate for soils outdoors and dust indoors for teens is half that of children, resulting in a 93rd percentile of 100 mg/day. This value is reported in Table 7. For the Monte Carlo simulations of teens, we divide the lognormal distribution for soil ingestion rate of children by a factor of two.

For adults, LaGoy (1987) notes that soil ingestion rates for adults who are frequently in contact with soil are about half those for older children. We assumed the ingestion rates for soils outdoors and dust indoors for adults to be described by the same distribution as that used for teens. The 93rd percentile of this distribution is 100 mg/day. This value is reported in Table 7. For the Monte Carlo simulations of adults, we divide the lognormal distribution for soil ingestion rate of children by a factor of two.

Region V: For default RME conditions, Region V assumes these rates for the incidental ingestion of soils outdoors: 200 mg/day for children, 200 mg/day for teens, and 100 mg/day for adults. Region V chooses these values for default RME conditions from the default values published among the Interim Final Standard Exposure Factors (US EPA, 1991, Default) for the Superfund program nationwide.

#### 6.3.3.2 Ingestion of Indoor Dust

Celotex and AlliedSignal: For RME conditions, Celotex and AlliedSignal assume that children, teens, and adults incidentally ingest 200, 100, and 100 mg/day of indoor dust on a day of indoor exposure. In the Monte Carlo simulations, Celotex and AlliedSignal assume the lognormal distributions listed in Table 7 for children, teens, and adults, as based on Thompson & Burmaster (1991). These RME values and distributions for ingestion of indoor dust are identical to those assumed for the ingestion of soils outdoors.

Region V: Since Region V assumes that all people ingest some soils outdoors for 350 days per year for default RME conditions, it is not applicable to have an exposure variable for the ingestion of dust indoors for default RME conditions. As mentioned earlier, Region V bases its default RME exposure scenario on the default values published among the Interim Final Standard Exposure Factors (US EPA, 1991, Default) for the Superfund program nationwide.

### 6.3.4 Exposure Point Concentration for Dust Indoors

Table 7 shows the assumptions made by Celotex and AlliedSignal and by Region V about the exposure point concentration for indoor dust.

Celotex and AlliedSignal: For RME conditions, Celotex and AlliedSignal use 0.42 (42 percent) as the transfer coefficient for the fraction of outdoor soil contributing to indoor dust. For the Monte Carlo simulations, Celotex and AlliedSignal use the lognormal distribution listed in Table 7 and developed in Appendix F as the full-information exposure factor. The point value (42 percent) is the median of the lognormal distribution.

As shown in Appendix F, dust inside a house contains some materials generated indoors and some materials carried into the house from outdoors. Using naturally-occurring conservative tracer chemicals, many researchers have measured the fraction of materials in house dust that originate outside a house. After analyzing these studies in detail and fitting a parametric distribution to them, we use 0.42 as the median "Transfer Coefficient" for this risk assessment. In other words, 42 percent of the material in house dust comes from soils outside the houses and the remainder comes from activities inside the house. Thus, for RME conditions, we model the exposure point concentration for BaPeq in dust inside the house as 42 percent of the concentration of

BaPeq in soils outside the house. As shown in Appendix F, we use a truncated lognormal distribution to model the Transfer Coefficient, i.e., the fraction of indoor dust that originates from outdoor soils. As shown in Table 7, the median of this distribution is 0.42 and the 95th percentile is 0.74 (see also Appendix F; Trowbridge & Burmaster, 1996).

Region V: For default RME conditions, Region V considers that all exposures occur outside the house (or equivalently, that all indoor exposures have a Transfer Coefficient equal to one so that concentrations inside the house are identical to the concentrations outside the house). As mentioned earlier, Region V bases its default RME exposure scenario on the default values published among the Interim Final Standard Exposure Factors (US EPA, 1991, Default) for the Superfund program nationwide.

### 6.3.5 Oral Absorption Adjustment Factor (AAF)

Table 7 shows the assumptions made by Celotex and AlliedSignal and by Region V for the oral absorption adjustment factor (oral AAF) for PAHs in soil and dust.

Exposure dose does not take into account the body's greater or lesser absorption of chemicals encountered in different media. We incorporated absorption adjustment factors (AAFs) into the risk equations to account for the difference between the measured concentration in the medium of exposure and the amount absorbed by the body. Celotex, AlliedSignal, and Region V all define the AAF as the ratio of the absorption (bioavailability) by the route and medium of interest to the absorption by the route and medium used in the dose-response study for the compound.

Celotex and AlliedSignal: For RME conditions, Celotex and AlliedSignal use 0.27 (27 percent) as the oral AAF from both outdoor soils and indoor dust. For the Monte Carlo simulations, Celotex and AlliedSignal use the four-parameter beta distribution listed in Table 7 and developed in Appendix F as the full-information exposure factor. The point value (27 percent) is the median of the four-parameter beta distribution.

Evidence in the literature suggests that PAHs adsorbed to soils are far less than 100 percent available when ingested. Thus, the dose received by an exposed individual comes from <100 percent of the PAH concentration in soil.

The Gas Research Institute funded a study of the behavior and toxicity of organic chemicals in soils (GRI, 1995). As part of this study, Prof. Martin Alexander of Cornell

University reviewed the literature to evaluate the current understanding of the bioavailability of organic compounds that have been present in soils for extended periods.

From his review, Alexander finds that organic compounds become more tightly bound to soil with increasing time because the compounds become sequestered within the soil itself, instead of simply remaining on the surface (GRI, 1995; Hatzinger & Alexander, 1995). This aging and sequestration affects the behavior of the compound bound to the soil, making the compounds much less available to biological systems (Bonaccorsi et al, 1984). This decreased extractability and bioavailability causes decreased biodegradability and toxicity. For chemicals like BaP that form strong bonds with soil, it is sometimes impossible to release the PAHs without using a Soxhlet extraction (boiling in strong acid for several hours). Alexander also indicates that bioavailability may also depend on the organic fraction of the soil.

As shown in Appendix G, Brian H. Magee, a toxicologist with Ogden Environmental and Energy Services Company, has quantified the oral AAF after reviewing the literature on the absorption of PAHs from foods and soils (Ogden, 1996). Based on the materials in Appendix G, we estimate the RME oral AAF from soil and dust as 0.29 (or 29 percent) for all three age groups. Based on the same information in Appendix G, we use a four-parameter beta distribution (Beta[1, 3, 0.945, 0.07]) in the Monte Carlo simulations.

Region V: The Region assumes that the oral AAF for soils equals 0.9 (or 90 percent) for default RME conditions for all three age groups considered in the risk assessment (Podowski, 1996). Although Region V does not explicitly consider the oral AAF for indoor dust, it is implicitly the same as the oral AAF for outdoor soils .

### 6.3.6 Adherence of Soil and Dust to Skin

Table 7 shows the assumptions for dermal exposure made by Celotex and AlliedSignal and by Region V about the adherence of soil and dust to the skin of children, teens, and adults.

#### 6.3.6.1 Adherence of Outdoor Soil to Skin

Celotex and AlliedSignal: Drawing on US EPA's published report (US EPA, 1992, Dermal), we assume that 1 mg/(cm$^2$•day) of outdoor soils adheres to the skin of

children, teens, and adults for RME conditions. This is greater than the 95th percentile of the full-information distribution developed from US EPA's data.

We find that a lognormal distribution gives a good fit to data published by the Agency (US EPA, 1992, Dermal) for the adherence of outdoor soils to skin of children, teens, and adults. The distribution -- exp[Normal[-1.71, 1.01)] -- has a median value of 0.18 mg/(cm2•day) and an upper 95th percentile of 0.95 mg/(cm2•day), a value in keeping with the RME point value.

Region V: Region V assumes the same amount of outdoor soil and indoor dust adhere to the skin of children, teens, and adults: 1 mg/($cm^2$•day) for default RME conditions. This number is the default value from Agency guidance (US EPA, 1992, Dermal).

#### 6.3.6.2 Adherence of Indoor Dust to Skin

Celotex and AlliedSignal: Based on the research, Kissel et al. (1996) report that less indoor dust adheres to skin than does outdoor soil because indoor dust has less moisture content than does outdoor soil. Based on this recently published research from the University of Washington, we assume that 0.2 mg/($cm^2$•day) of indoor dust adheres to the skin of children, teens, and adults for RME conditions (Kissel et al, 1996). By dividing the distribution for the adherence rate of outdoor soil to skin by a factor of five, we arrive at the distribution for the adherence rate of indoor dust to skin. This lognormal distribution has a median of 0.05 and a 95th percentile of 0.24 (see Table 7).

Region V: Since Region V assumes that all people have dermal contact with soils outdoors for 350 days per year for default RME conditions, it is not applicable to have an exposure variable for the dermal contact with dust indoors for default RME conditions. To the extent that the Region allows for exposure to indoor dust, the adherence rate for indoor dust for default RME conditions is the same as the adherence rate for outdoor soils for default RME conditions. As before, Region V bases its RME exposure scenario on the default values published among the Interim Final Standard Exposure Factors (US EPA, 1991, Default) for the Superfund program nationwide.

### 6.3.7 Total Skin Surface Area

Table 7 shows the assumptions made by Celotex and AlliedSignal and by Region V about the total skin surface area for children, teens, and adults.

Celotex and AlliedSignal: For RME conditions, we assume these values for total skin surface area from US EPA guidance documents: 0.73 $m^2$ for children, 1.5 $m^2$ for teens, and 2 $m^2$ for adults. (US EPA, 1990, EFH). In the probabilistic exposure assessment, we use Costeff's formula (Costeff, 1966; Murray & Burmaster, 1992) to estimate the total skin surface area of children, teens, and adults as a function of body weight. Although Costeff originally developed his formula for children and teens, it also provides a good fit for adults as well.

Region V: Region V assumes the same total skin surface area for RME conditions: 0.73 $m^2$ for children, 1.5 $m^2$ for teens, and 2 $m^2$ for adults. These numbers are consistent with default values presented in Agency guidance (e.g., US EPA, 1990, EFH).

### 6.3.8 Fraction of Skin Exposed

Table 7 shows the values assumed by Celotex and AlliedSignal and by Region V to represent the fraction of skin exposed to dermal contact with outdoor soils and indoor dust by the different age groups.

Celotex and AlliedSignal: Celotex and AlliedSignal assume that each child, teen, and adult has 25 percent of his or her total skin area exposed to dermal contact to soils outdoors for RME conditions. According to tables published by US EPA (Anderson, 1984), 25 percent coverage corresponds, for example, to having dermal contact with the forehead, face, both ears, neck, both hands, both forearms, and both feet everyday. According to the same tables, 25 percent coverage also corresponds to having dermal contact with both hands, both feet, and both lower legs. In the probabilistic exposure assessment, Celotex and AlliedSignal also use 25 percent as a point estimate for the fraction of skin exposed.

Region V: Region V assumes that each person has 25 percent of his or her total skin area exposed to dermal contact to soils outdoors for default RME conditions. Although Region V does not explicitly consider the fraction of skin exposed for indoor dust, it is implicitly the same as the fraction of skin exposed for outdoor soils.

### 6.3.9 Dermal Absorption Adjustment Factor (AAF)

Table 7 shows the assumptions made by Celotex and AlliedSignal and by Region V about the dermal absorption adjustment factor for soil and dust.

Exposure dose does not take into account the body's greater or lesser absorption of chemicals encountered in different media. In estimating potential health effects, the medium of exposure for which a dose is being estimated must be the same as the medium of exposure on which the published toxicity value is based. Celotex, AlliedSignal, and Region V all incorporate absorption adjustment factors (AAFs) into the risk equations to account for the difference in absorption of the applied dose in the medium of exposure. The AAF is defined as the ratio of absorption (bioavailability) by the route and medium of interest to absorption by the route and medium used in the dose-response study for the compound.

Celotex and AlliedSignal: As shown in Appendix G, Brian H. Magee has quantified the dermal AAF (Ogden, 1996). Based on Appendix G, we estimate the dermal AAF for outdoor soils and indoor dust as 0.02 (or 2 percent) for RME conditions for all three age groups. Based on Appendix G, we model the dermal AAF as the ratio of two scaled and translated beta distributions, where the numerator equals Beta4[1, 5, 0.147, 0] and the denominator equals Beta4[4, 1, 0.397, 0.603] (see Table 7, Appendix G, and Burmaster, 1996).

Region V: Region V assumes that the dermal AAF for outdoor soils equals 0.15 (or 15 percent) for default RME conditions for all three age groups (Podowski, 1996).

### 6.3.10 Other Exposure Assumptions

Table 7 also presents the values assumed for other exposure variables. The values chosen by Celotex and AlliedSignal and by US EPA Region V are in keeping with generally accepted values in risk assessment for the Superfund program (US EPA, 1989, HHEM; US EPA, 1990, EFH).

#### 6.3.10.1 Lifetime

As shown Table 7, Celotex, AlliedSignal, and Region V assume that a person lives 70 years for both the deterministic and the probabilistic calculations.

#### 6.3.10.2 Body Weight

Celotex, AlliedSignal, and Region V make similar assumptions regarding the body weight of children, teens, and adults. Table 7 shows the values for the three age groups.

Celotex and AlliedSignal: For children for RME conditions, we average reported annual average body weights from age 1 year through age 6 years (Anderson et al., 1984). The average body weight for boys and girls in this age group is 16.4 kg. For teens for RME conditions, we average reported annual average body weights from age 7 years through age 17 years (Anderson et al., 1984). The average body weight for boys and girls in this age group is 44.9 kg. For adults for RME conditions, we use the standard US EPA value for the average weight of an adult men and women, 70 kg.

For the simulations, we developed lognormal distributions for the body weights of children and teenagers (using equal proportions of boys and girls in each age group) based on fitted distributions in the literature (Burmaster, Lloyd & Crouch, 1994). For children, we use the lognormal distribution exp[Normal(2.69, 0.33)] for body weights measured in kg. The median, mean, and 90th percentile of this distribution are 14.7, 15.6, and 22.5 kg, respectively. For teens, we use the lognormal distribution exp[Normal(3.75, 0.37)] for body weights measured in kg. The median, mean, and 90th percentile of this distribution are 42.5, 45.5, and 68.3 kg, respectively. We use this published lognormal distribution for the body weights of adults measured in kg: exp[Normal(4,263m 0.206)] (Table 7; and Brainard & Burmaster, 1992).

Region V: Region V uses 15, 45 and 70 kg as the RME body weights for children, teens, and adults, respectively. The values come from the default values published among the Interim Final Standard Exposure Factors (US EPA, 1991, Default) for the Superfund program nationwide or from other Agency tables (e.g., US EPA, 1990, EFH).

## 6.4 Estimation of Doses to Populations of Concern

We used the equations shown in Table 9 to estimate doses and cancer risk. Table 10 shows the formulas used in the spreadsheet. The methods used to calculate the average daily doses are described in the following subsections.

### 6.4.1 Forward Calculation of RBCGs

In their deterministic risk assessment, E&E (1995) calculated point values for RBCGs by rearranging the risk equation, substituting the single target cancer risk, and solving for the fixed exposure point concentration. We refer to this calculation as a backward calculation. The backward calculation is acceptable in deterministic RBCGs when all the variables are point values. However, this backward calculation fails when using

probability distributions to calculate the RBCG as a distribution (Burmaster, Lloyd & Thompson, 1995; Burmaster & Thompson, 1995).

An alternative calculation, one that gives correct answers for all calculations, consists in calculating the acceptable exposure point concentration with the original risk equation (i.e., without using algebra to rearrange the risk equation). Since the RBCG is not solved for directly, it must be solved iteratively, by substituting different values for the RBCG until the target cancer risk (or a value slightly smaller) is reached. We refer to this direct approach as the iterative forward calculation. This iterative forward method works correctly for both deterministic and probabilistic calculations (Burmaster, Lloyd & Thompson, 1995; Burmaster & Thompson, 1995).

### 6.4.2 Estimation of Average Daily Dose on a Day of Exposure

For each Study Chemical, we use the equations in Table 9 to estimate an average daily dose on a day of exposure, ADD(day), separately for each exposure pathway and each life stage that contribute to a given scenario. All ADD(day) values are in units of milligrams of Study Chemical per kilogram body weight per day (mg/(kg•day)).

As shown in the formulae in Table 9, the ADD(day) is calculated for each age group by multiplying the Acceptable Exposure Point Concentration, the appropriate Absorption Adjustment Factor, the Contact Rate for the age group, and the Conversion Factor and by dividing by the Body Weight for the age group.

### 6.4.3 Estimation of Average Daily Dose over a Lifetime

For each Study Chemical, we use Equation 3 in Table 9 to estimate an average daily dose over a lifetime of exposure, ADD(life). The ADD(life) takes into account the frequency (days per year) with which exposure occurs.

We estimate ADD(life) separately for each age group by multiplying the average daily dose on a day of exposure (ADD(day)) by the Exposure Frequency and an appropriate Conversion Factor.

### 6.4.4 Estimation of Average Daily Dose (Total)

We use Equation 3 in Table 9 to estimate a total average daily dose over a lifetime of exposure, ADD(total). The total average daily dose experienced during a lifetime,

ADD(total), takes into account the fraction of a lifetime during which the exposure occurs. Thus, ADD(total) is equal to ADD(life) if exposure occurs throughout the lifetime, but is smaller than ADD(life) if exposure occurs during only some years. Of course, the total ADD is the sum of the doses received from the ingestion pathway and the dermal contact pathway.

We estimate ADD(total) by multiplying the value of ADD(life) for each age group by the duration of exposure for that age group, summing the three resulting products (one for each age group), and dividing by the number of years in a lifetime.

# 7. Risk Characterization

## 7.1 Selection of Target Cancer Risk

### 7.1.1 Target Risk for Deterministic Calculations

In their earlier deterministic risk assessments, both E&E (1995) and Alceon (1996) used the target cancer risk of $10^{-4}$ (or one in ten thousand).

This risk management target is consistent with the current National Contingency Plan (US EPA, 1990, NCP), published in the Federal Register in March 1990:

> For known or suspected carcinogens, acceptable exposure levels are generally concentration levels that represent an excess upper bound lifetime cancer risk to an individual of between $10^{-4}$ and $10^{-6}$ using information on the relationship between dose and response. The $10^{-6}$ risk level shall be used as the point of departure for determining remediation goals for alternatives when ARARs are not available or are not sufficiently protective because of the presence of multiple contaminants at a site or multiple pathways of exposure. (p. 8848)

The "point of departure" reflects US EPA's preference for remedies that are at the more protective end of the risk range. However,

> a variety of site-specific or remedy-specific factors . . . will enter into the determination of where within the risk range of $10^{-4}$ to $10^{-6}$ the cleanup standard for a given contaminant will be established. (p. 8717)

These factors may be related prevailing background concentrations, to other exposures to the same compounds in foods or other media, to other exposure issues (e.g., the cumulative effect of multiple contaminants), to uncertainty (e.g., the weight of scientific evidence concerning health effects), or to technical issues (e.g., detection limits for contaminants).

### 7.1.2 Target Risk for Probabilistic Calculations

Following these precedents and the concepts presented in US EPA (1992, Exposure), we judge the acceptability or unacceptability of a distribution of risk using two constraints on the distribution (Burmaster & Thompson, 1994).

- For a first constraint -- and in keeping with the approach taken by E&E -- we associate a target cancer risk of 1 in 10,000 (equivalent to $10^{-4}$) with the 95th percentile of the risk distribution. In other words, for a distribution of risk to be acceptable, its 95th percentile must be $\leq 10^{-4}$. In US EPA's terminology, this constraint takes the place of the Reasonable Maximum Exposure (RME exposure).

AND

- For a second constraint, we associate a target cancer risk of 1 in 100,000 (equivalent to $10^{-5}$ risk) with the median risk. In other words, for a distribution of risk to be acceptable, its 50th percentile (median) must be $\leq 10^{-5}$. In US EPA's terminology, this constraint takes the place of the Central Tendency Exposure (CT exposure).

To be acceptable under this definition, a distribution of risk must meet both the constraints simultaneously. These two simultaneous constraints create a risk management policy that is more stringent than the one used by Ecology & Environment.

These constraints agree with the concepts of (i) "high-end" risk or "Reasonable Maximum Exposure" risk and (ii) "typical" risk or "Central Tendency" risk as used in US EPA guidance in the Federal Register (US EPA, 1992, Exposure). This risk management policy -- based on two simultaneously binding constraints, one on the 95th percentile risk and one on the median risk -- resembles policies recently accepted by the US EPA at various Superfund sites when remediating or decommissioning facilities run by the US Department of Defense (US DOD) or by the US Department of Energy (US DOE).

## 7.2 Estimation of RBCGs

Tables 9 and 10 show the formulae used to estimate the incremental lifetime cancer risk (ILCR) from the total average daily dose. In keeping with US EPA's methods to estimate the pathway-specific ILCRs (US EPA, 1989, HHEM), we multiply the ADD(total) for each

Study Chemical by the appropriate CSF (i.e., the ingestion or inhalation CSF for that Study Chemical).

Table 10 shows the formulae in the spreadsheet that were used to calculate a target cancer risk of just equal to or less than $10^{-4}$ risk. This target cancer risk is enclosed by a box and is located in the lower right-hand corner of the spreadsheet.

## 7.3 Estimated RBCGs for the Study Area

### 7.3.1 RBCG Based on Celotex's and AlliedSignal's Deterministic Assumptions

Using Celotex's and AlliedSignal's exposure factors for RME conditions in Table 7, we calculate that the deterministic RBCG for BaPeq in outdoor surface soils in the residential neighborhoods near the industrial property equals 27.5 mg/kg BaPeq for RME conditions (see calculations in Table 11).

### 7.3.2 RBCG Based on Celotex's and AlliedSignal's Probabilistic Assumptions

Table 13 shows the report from Crystal Ball® (Decisioneering, 1992). From the full simulation of 20,000 iterations and from the report, we see that an acceptable distribution of risk (one that meets both constraints defining the acceptability of a distribution of total risk for both the ingestion and the dermal pathways) occurs when the BaPeq concentration follows the distribution below. Since this distribution of BaPeq concentrations causes a distribution of risk that simultaneously meets (read, is equal to or less than) the two constraints defining the maximum acceptable risk, it is, by definition, the cleanup target for BaP concentration in soils.

| | | | |
|---|---|---|---|
| minimum | = | 0 | mg/kg BaPeq |
| 10th percentile | ≤ | 6.4 | mg/kg BaPeq |
| 20th percentile | ≤ | 9.5 | mg/kg BaPeq |
| 30th percentile | ≤ | 12.8 | mg/kg BaPeq |
| 40th percentile | ≤ | 16.4 | mg/kg BaPeq |
| 50th percentile | ≤ | 20.8 | mg/kg BaPeq |
| 60th percentile | ≤ | 26.1 | mg/kg BaPeq |

| | | | |
|---|---|---|---|
| 70th percentile | ≤ | 33.3 | mg/kg BaPeq |
| 80th percentile | ≤ | 43.6 | mg/kg BaPeq |
| 90th percentile | ≤ | 58.2 | mg/kg BaPeq |
| 95th percentile | ≤ | 72.7 | mg/kg BaPeq |
| maximum | ≤ | 99.9 | mg/kg BaPeq |

This distribution, a lognormal distribution truncated at 100 mg/kg BaPeq, has an arithmetic mean equal to 27.1 mg/kg BaPeq. See Figure 2. In a statistical sense, this probabilistic RBCG dominates the distribution of background concentrations discussed earlier (Clemen, 1991; see also Appendices A and B). In a colloquial sense, this probabilistic RBCG "is larger than" the distribution of background concentrations in Appendix A.

From the simulation with 20,000 iterations and the report from Crystal Ball® (Decisioneering, 1992), we verify that the Incremental Lifetime Cancer Risk attributable to this distribution of BaPeq in soil meets the two constraints that define the acceptability of a distribution of total risk (the sum of the risk from the ingestion and dermal pathways):

| | | | | | |
|---|---|---|---|---|---|
| | 10th percentile | = | $10^{-6.28}$ | | |
| | 20th percentile | = | $10^{-5.88}$ | | |
| | 30th percentile | = | $10^{-5.61}$ | | |
| | 40th percentile | = | $10^{-5.39}$ | | |
| --> | 50th percentile | = | $10^{-5.19}$ | ≤ | $10^{-5}$ |
| | 60th percentile | = | $10^{-5.00}$ | | |
| | 70th percentile | = | $10^{-4.80}$ | | |
| | 80th percentile | = | $10^{-4.58}$ | | |
| | 90th percentile | = | $10^{-4.30}$ | | |
| --> | 95th percentile | = | $10^{-4.08}$ | ≤ | $10^{-4}$ |

The arithmetic average of this distribution of Incremental Lifetime Cancer Risk equals $10^{-5.24} = 5.75 \bullet 10^{-6}$ risk, well within the policy in the National Contingency Plan (US EPA, 1990, NCP). Interestingly, the risk distribution from the ingestion pathway dominates the risk distribution from the dermal pathway.

#### 7.3.3 RBCG Based on Region V's Deterministic Assumptions

Using US EPA Region V's default exposure factors for RME conditions in Table 7, we calculate that the deterministic RBCG for BaPeq in outdoor surface soils in the residential neighborhoods near the industrial property equal 1.93 mg/kg BaPeq for default RME conditions (see calculations in Table 12).

### 7.4 Interpretation and Application of the RBCGs

Under the Superfund statute, a RBCG is commonly calculated in the risk assessment portion of a Remedial Investigation (the RI), while the engineering method(s) of achieving that RBCG are developed in the Feasibility Study (the FS). In the Record of Decision (ROD), the risk manager has considerable latitude in selecting the best risk management option for a situation, taking into account all the goals, policies, desiderata, and balancing criteria -- including background concentrations, engineering feasibility, economics, and public acceptability -- in the National Contingency Plan (US EPA, 1990, NCP).

No matter how the risk manager arrives at the risk management intervention during the FS study and in the ROD, he or she must consider four fundamental issues that we have not heretofore considered in this report:

- a RBCG for surface soils has the character of an Exposure Point Concentration (EPC) for those surface soils,
- an EPC is defined by the US EPA as an arithmetic mean concentration over the variability in a set of measurements (or often as the 95th percentile of the uncertainty in the arithmetic mean concentration),
- EPCs for soils must consider the two-dimensional spatial character of the concentrations in surface soils before and after the proposed remediation, and
- EPCs, as spatial average values, apply over a population of spatial measurements where a population of people have exposures.

Therefore, regardless of whether the RBCG is calculated using deterministic or probabilistic methods, an RBCG is not the maximum allowable concentration of BaPeq in surface soils at every single point in space, nor is it the maximum allowable concentration of BaPeq experienced by any single person.

Even in a situation where the risk manager has chosen a point value as the RBCG for surface soils, an EPC fully meeting that RBCG over an area will necessarily contain some or many individual concentration measurements at spatial points that exceed the point value of the RBCG. Similarly, in a population of people in an area that meets an EPC achieving a point value RBCG, there will be some individuals who have higher personal EPCs than do other individuals living or working in the same area. These facts arise from the fundamental and inalienable two-dimensional spatial nature of concentrations in surface soils. Succinctly, a RBCG does not apply to the maximum single concentration in an area, nor does it apply to the maximally exposed person.

Alternately, in a situation where the risk manager has chosen a distribution as the RBCG for surface soils, a distribution of EPCs fully meeting that distributional RBCG may well contain some or many individual concentration measurements at spatial points that seem to fall outside the distributional RBCG yet achieve the EPC.

Of course, based on US EPA's national policy, it is not appropriate to have a RBCG that is lower than the natural or anthropogenic background concentrations for the same compounds in the same media in a similar but unaffected area. Celotex and AlliedSignal understand that any risk management decision based on treating one area to match the background concentrations in a nearby but unaffected area (taken as the background area) must consider both (i) matching the spatial patterns of concentrations in the treatment area and the background area and (ii) matching the distributions of concentrations in the treatment area and the background area.

For the neighborhoods surrounding the industrial property, it is premature to design or even conceptualize the engineering methods to meet a particular risk management decision -- whether that decision is couched in terms of meeting a distribution of background concentrations or in terms of meeting a deterministic or probabilistic RBCG. While we cannot resolve all these issues in this report, we can describe their operation in broad brush.

### 7.4.1 Interpretation of a Deterministic RBCG

Since we have calculated the RBCG in the forward direction for RME and default RME sets of assumptions, each of these alternate deterministic RBCGs has the character and interpretation of an Exposure Point Concentration (EPC) for outdoor surface soils. When interpreting a deterministic RBCG for outdoor surface soils, the US EPA recommends calculating the EPC as the 95th-percentile upper confidence level (UCL) of the uncertainty on the arithmetic mean concentration of variability experienced by the exposed population US EPA, 1992, EPC. This EPC for outdoor surface soils is not the maximum concentration in an area, and it is not calculated for each single property, one at a time.

Therefore, to interpret or apply one of the RBCGs for outdoor surface soils in a residential area near the industrial property, we would not consider just one property at a time. In keeping with US EPA's guidance on developing EPCs for surface soils, we would instead consider the 95th percentile upper confidence limit (UCL) on the arithmetic mean of the surface soil concentrations averaged over the many properties where each person has exposure (US EPA, 1992, EPC). Since no person has all of his or her exposure on a single property, the EPC is properly calculated over many properties or over the whole neighborhood using activity-, time-, and distance-weighted spatial statistics.

### 7.4.2 Interpretation of a Probabilistic RBCG

Since we have calculated this probabilistic RBCG in the forward direction, it also has the character and interpretation of a distribution of Exposure Point Concentrations (EPCs) for outdoor soils. Again, since no person has all of his or her exposure on a single property, the EPC is properly calculated over many properties or over the whole neighborhood using activity-, time-, and distance-weighted spatial statistics. To apply this probabilistic cleanup target in practice, remedial engineers would compare this distribution for the RBCG directly to the measured distribution of BaPeq concentrations in the surface soils in the population of yards in the residential neighborhoods in the Study Area near the industrial property. Two cases might arise:

- First, the distribution of EPCs for measured BaPeq concentrations in surface soils may have percentiles and a maximum that all fall below the corresponding percentiles and below the maximum (100 mg/kg) for BaPeq in the RBCG. In this case, no remedial action is necessary or appropriate.

- Second, the distribution of EPCs for measured BaPeq concentrations in surface soils may have certain percentiles or a maximum that exceeds the corresponding percentiles or maximum (100 mg/kg) for BaP in the RBCG. In this case, some type of remediation may be considered.

If remediation is considered, there are two bedrock principles that should guide the program:

- The "Worst First" Principle - As enunciated by Resources for the Future (Finkel & Golding, 1994) for situations like this, the "Worst First" policy states that the optimal public health policy includes two steps:

    1. prioritize the problems and opportunities, and

    2. focus resources on the worst of the problems first.

- The "Distribution Matching" Principle -- As stated in texts on decision science (e.g., Clemen, 1991), it is both possible and desirable from both an efficiency and equity points of view to undertake interventions to make the distribution of field conditions match the distribution of the goal.

In practice, a remedial engineer might implement these two policies along these lines. The engineer would implement this algorithm (or one like it) in an iterative fashion, searching for an optimal solution, i.e., a solution that meets the risk management policy in the most cost-effective way:

- First, the engineer would measure the BaPeq concentrations in the surface soils in all the residential yards in the neighborhoods surrounding the industrial property.

- Second, the engineer would compute the activity-, time-, and distance-weighted spatial statistics and estimate the distribution of EPCs.

- Third, the engineer would compare the distribution of EPCs based on measured concentrations to the distribution for the RBCG. If any percentiles of the EPC distribution exceed the corresponding percentile or maximum of the probabilistic RBCG, then the engineer would remediate the surface soils

at one or more of the more highly contaminated properties (say, to a concentration below the background distribution).

- If the distribution of the EPCs based on the remaining measured concentrations still does not meet the distribution for the RBCG, then the engineer would remediate the surface soils at other highly contaminated properties.
- The engineer would continue to remediate the surface soils at the highly contaminated properties remaining on the list until the distribution of EPCs based on the remaining measured concentrations in the surface soils no longer has percentiles that do not meet the percentiles of the RBCG distribution.
- When this algorithm stops, the distribution of EPCs in the neighborhoods will be statistically smaller than the RBCG distribution, i.e., the distribution of EPCs will meet all the constraints defined by the RBCG distribution.

An engineer would use a similar algorithm to achieve a cleanup in an area specified in terms of achieving background concentrations.

# 8. Uncertainty Analyses

## 8.1 General Discussion for All Deterministic and Probabilistic Calculations

As in any risk assessment, the factors considered in this report contain both variability and uncertainty. As we use the terms, variability represents heterogeneity or diversity in a well-characterized population, usually not reducible through further measurement or study. Uncertainty represents ignorance -- or lack of perfect information -- about a poorly-characterized phenomenon or model, sometimes reducible through further measurement or study.

Appendix H examines uncertainties in the exposure and health effects data that are relevant for assessing potential human health risks associated with exposure to contaminants originating from the industrial property. Many of the key quantities considered in the risk assessment are highly uncertain. For example, the following factors have not been estimated with high precision or confidence:

- The spatial distribution and extent of contamination from the industrial property in various directions is not well known. Instead, it must be estimated from soil sample data. (See discussion in Appendix B).
- The fraction of PAHs found at any specific location that arise from the industrial property is uncertain. The problem of distinguishing between site-related and "background" contamination arises, since the same contaminants and approximate composition of PAHs found near the industrial property are also found at distances remote enough to make association with the industrial property implausible.
- The magnitudes and frequencies of individual exposures depend on individual behaviors and on details of the yards (e.g., extent of vegetative cover as opposed to rock and debris cover) that have not been quantified. Hence, the actual magnitude of individual exposures is uncertain. Drive-by inspection of yards in the vicinity of the industrial property suggests that they are dissimilar in many respects (e.g., more rubble, less accessible soil useful for gardening or recreation) compared to locations further from the industrial property. How these local characteristics affect individual behaviors and exposures has not been estimated. Similarly, local demographic characteristics (e.g., the ages, occupations, recreational patterns, etc.) of neighbors of the industrial property have not been examined. Yet, these characteristics may affect the magnitudes and frequencies of individual exposures to yard soils.
- The amounts of internal doses of reactive, potentially carcinogenic PAH metabolites formed in humans at the exposure levels in question are not

known. In particular, the relative amounts of internal doses formed in humans compared to the amounts formed in animals under the experimental conditions used to establish the carcinogenicity of PAHs such as B(a)P are not known.

- The cancer potency of PAHs, including B(a)P, at the concentrations found near the industrial property is not known. Specifically, the relation between carcinogenic potency of B(a)P at the high doses used in animal carcinogenicity experiments and its potency at the much lower levels found in the soil samples examined in this study is not known. In addition, the potency of the PAH mixtures found in the soil samples is uncertain.

These uncertainties create a challenge for fair, efficient, health-protective risk management. The actual human health risks posed by the industrial property are not known. They would be costly to quantify with high precision and confidence, since doing so would require resolving each of these sources of uncertainty. Yet, it is desirable to avoid the two types of risk management errors most likely to occur in this case: failure to adequately reduce site-related exposures, and failure to limit reductions to those that significantly reduce actual human health risks. The purpose of the analyses reported in this appendix and the next one is to reduce the probabilities of both types of errors by introducing relevant information and findings from recently completed data analyses and literature reviews. A suggested approach to risk management decision-making in the presence of the uncertainties just listed is offered after some relevant facts, data, and statistical results have been summarized.

Appendix B focuses on the first issue -- uncertainty about the spatial extent of contamination near the industrial property. This issue can be addressed without considering risk assessment questions and data: it rests solely on statistical analysis of the soil sample data collected so far. Analysis of these data reveals the maximum probable geographic extent of effects from the industrial property, and thus provides a basis for bounding the geographic scope of the risk assessment without regard for risk magnitudes. Appendix H presents the remaining sources of uncertainty and their implications for risk management.

## 8.2 Semi-Quantitative Uncertainty Analysis for Deterministic Calculations

Both sets of deterministic calculations in this report include compounding conservatisms in the sense that each calculation includes some conservative assumptions and some median or average assumptions for exposure variables. When a calculation is done

using only median values, the calculation does not contain compounding conservatisms because medians are statistically neutral in multiplicative equations (Burmaster & Bloomfield, 1996). However, calculations involving average values and other percentiles greater than the 50th percentile do create compounding conservatisms in the risk equation (Burmaster & Bloomfield, 1996)

### 8.2.1 Celotex's and AlliedSignal's RME Assumptions

In developing the inputs for the Celotex and AlliedSignal risk assessment, our general approach to uncertainty has been to use an appropriate combination of health-protective assumptions in estimating exposures, so that the cancer risks that we estimate are based on a "high end exposure" to compensate for the uncertainties inherent in this analysis (as defined in US EPA, 1992, Exposure). We believe that we have included exposure assumptions that are reasonable for the Study Area and which consider sensitive sub-populations, especially children and teens.

For both the ingestion pathway and the dermal pathway for children, teens, and adults, we use a balanced combination of median values (which are statistically neutral), average values (which introduce a moderate amount of statistical compounding), and some conservative values (≥90th percentiles; which introduce a strong amount of statistical compounding) in the exposure calculations.

Median Values (50th Percentiles)
- Oral Absorption Adjustment Factor
- Dermal Absorption Adjustment Factor
- Exposure Frequencies Outdoors and Indoors

Average Values (between 50th and 85th Percentiles)
- Transfer Coefficient to Indoor Dust
- Body Weight
- Skin Area
- Fraction of Skin Exposed

Conservative Values (≥90th Percentiles)
- Soil and Dust Ingestion Rates
- Exposure Duration

Soil and Dust Adherence Rates

The conservative factors alone -- not counting the compounding effects of the average values -- compound to create a value of exposure above the 95th percentile of exposure (US EPA, 1992). When this high value is multiplied by the 95th percentile for the Cancer Slope Factor (from US EPA's IRIS database), the resulting estimate of risk has a still higher percentile. In addition, we have included other conservatisms, for example, by assuming that all exposures in the population begin at age 1 year.

### 8.2.2 US EPA Region V's Default RME Assumptions

On the other hand, the US EPA Region V has included many more strongly compounding conservatisms, giving the results with their assumptions the character of a "bounding estimate" (as defined in US EPA, 1992, Exposure).

For both the ingestion pathway and the dermal pathway for children, teens, and adults, US EPA Region V uses a different combination of values, with fewer median values (which are statistically neutral) and more conservative values (≥90th percentiles; which introduce a strong amount of statistical compounding) in the exposure calculations.

Median Values (50th Percentiles)
- None

Average Values (between 50th and 85th Percentiles)
- Transfer Coefficient to Indoor Dust
- Body Weight
- Skin Area
- Fraction of Skin Exposed

Conservative Values (≥90th percentile, some ≥95th percentile)
- Oral Absorption Adjustment Factor
- Dermal Absorption Adjustment Factor
- Soil and Dust Ingestion Rates
- Exposure Duration (≥95th percentile)
- Exposure Frequencies Outdoors and Indoors (≥95th percentile)
- Soil and Dust Adherence Rates

The conservative factors alone -- not counting the compounding effects of the average values -- compound to create a value of exposure far above the 95th percentile of exposure (US EPA, 1992). When this high value is multiplied by the 95th percentile for the Cancer Slope Factor (from US EPA's IRIS database), the resulting estimate of risk has a still higher percentile -- a value in the range of a "bounding estimate" (as defined in US EPA, 1992, Exposure).

## 9. Summary of the Risk Based Cleanup Goals for Surface Soils

Alceon performed a human health risk assessment that estimates different sets of risk-based cleanup goals for the concentration of BaPeq in surface soils at homes near the industrial property in Chicago. The RBCGs are based on acceptable exposure point concentrations of carcinogenic PAHs in soils in the vicinity of the industrial property.

Using deterministic methods -- techniques that are neither full-information nor state-of-the-art -- Alceon estimates the neighborhood-specific risk-based cleanup goals for surface soils outside the residential houses near the industrial property as follows:

| | | |
|---|---|---|
| Celotex's and AlliedSignal's RME Assumptions: | 27.5 | mg/kg BaPeq |
| US EPA Region V's Default RME Assumptions: | 1.93 | mg/kg BaPeq |

Using probabilistic methods -- techniques that are both full-information and state-of-the-art -- we see that an acceptable distribution of risk (one that meets both constraints defining the acceptability of a distribution of total risk for both the ingestion and the dermal pathways) occurs when the BaPeq concentration follows the distribution below. Since this distribution of BaPeq concentrations causes a distribution of risk that simultaneously meets (read, is less than) the two constraints defining the maximum acceptable risk, it is, by definition, the cleanup target for BaP concentration in soils.

| | | | |
|---|---|---|---|
| minimum | = | 0 | mg/kg BaPeq |
| 10th percentile | ≤ | 6.4 | mg/kg BaPeq |
| 20th percentile | ≤ | 9.5 | mg/kg BaPeq |
| 30th percentile | ≤ | 12.8 | mg/kg BaPeq |
| 40th percentile | ≤ | 16.4 | mg/kg BaPeq |
| 50th percentile | ≤ | 20.8 | mg/kg BaPeq |
| 60th percentile | ≤ | 26.1 | mg/kg BaPeq |
| 70th percentile | ≤ | 33.3 | mg/kg BaPeq |

| | | | |
|---|---|---|---|
| 80th percentile | ≤ | 43.6 | mg/kg BaPeq |
| 90th percentile | ≤ | 58.2 | mg/kg BaPeq |
| 95th percentile | ≤ | 72.7 | mg/kg BaPeq |
| maximum | ≤ | 99.9 | mg/kg BaPeq |

This distribution, a lognormal distribution truncated at 100 mg/kg BaPeq, has an arithmetic mean equal to 27.1 mg/kg BaPeq. In a statistical sense, this probabilistic RBCG dominates the distribution of background concentrations discussed earlier (Clemen, 1991; see also Appendices A and B). In a colloquial sense, this probabilistic RBCG "is larger than" the distribution of background concentrations in Appendix A.

## 10. Limitations

Alceon has used reasonable care in performing all of the analyses in this report. Alceon has performed its services based upon risk assessment practices accepted at the time they were performed.

# 11. Abbreviations and Acronyms

| | | |
|---|---|---|
| AAF | = | Absorption Adjustment Factor (oral or dermal) |
| ADD | = | Average Daily Dose |
| ADD(day) | = | Average Daily Dose averaged over a day on which exposure occurs |
| ADD(year) | = | Average Daily Dose averaged over a year on which exposure occurs |
| ADD(life) | = | Average Daily Dose averaged over a lifetime of 75 years |
| BaP | = | Benzo(a)pyrene |
| BaPeq | = | Benzo(a)pyrene Toxic Equivalents |
| BW | = | Body Weight |
| cPAHs | = | Carcinogenic Polycyclic Aromatic Hydrocarbons |
| CSF | = | Cancer Slope Factor |
| CT | = | Central Tendency |
| d | = | day |
| DL | = | Detection Limit |
| E & E | = | Ecology and Environment |
| EE/CA | = | Engineering Evaluation/Cost Analysis |
| EPC | = | Exposure Point Concentration |
| ERM | = | Environmental Resources Management |
| ft | = | feet |
| g | = | gram |
| hr | = | hour |
| H | = | Henry's Law Constant |
| HEAST | = | Health Effects Assessment Summary Tables (US EPA) |
| HEE | = | High End Exposure |
| HHEM | = | Human Health Evaluation Manual (US EPA) |
| HI | = | Hazard Index |
| IARC | = | International Agency for Research on Cancer |
| IEPA | = | Illinois Environmental Protection Agency |
| ILCR | = | Incremental Lifetime Cancer Risk |
| IRIS | = | Integrated Risk Information System (US EPA) |
| kg | = | kilogram |
| $K_{oc}$ | = | Partition coefficient between water and organic carbon |
| $K_{ow}$ | = | Partition coefficient between water and octanol |
| Kp | = | Permeability Coefficient |
| l | = | liter |
| MCL | = | Maximum Contaminant Level |
| MCLG | = | Maximum Contaminant Level Goal |
| m | = | meter |
| $m^3$ | = | cubic meter |
| mg | = | milligram |
| µg | = | microgram |
| mm | = | millimeter |
| mol | = | mole |
| MW | = | Molecular Weight |
| NCP | = | National Contingency Plan |
| ncPAHs | = | NonCarcinogenic Polycyclic Aromatic Hydrocarbons |
| NOAEL | = | No Observed Adverse Effect Level |
| OSHA | = | Occupational Safety and Health Administration |
| PAHs | = | Polycyclic Aromatic Hydrocarbons |
| PHRED | = | Public Health Risk Evaluation Database (US EPA) |
| PM10 | = | Concentrations of Particulate Matter Less Than 10µm in Diameter |
| PA | = | Preliminary Assessment |
| ppb | = | parts per billion |
| ppbv | = | parts per billion by volume |

| | | |
|---|---|---|
| PRP | = | Potentially Responsible Party |
| RBCG | = | Risk-Based Cleanup Goals |
| RfD | = | Reference Dose |
| RME | = | Reasonable Maximum Exposure |
| S | = | Solubility (aqueous) |
| SSI | = | Screening Site Inspection |
| SSP | = | Support Sampling Plan |
| TC | = | transfer coefficient |
| US EPA | = | US Environmental Protection Agency |
| Vp | = | vapor pressure |
| yr | = | year |

# 12. References

NB See also references in each Appendix

AIHC, 1994
American Industrial Health Council, 1994, Exposure Factors Source Book, Washington, DC

Aitchison & Brown, 1957
Aitchison, J. and J.A.C. Brown, 1957, The Lognormal Distribution, Cambridge University Press, Cambridge, UK

Alceon, 1996, DRA
Alceon Corporation, 1996, Deterministic Calculations to Estimate Risk-Based Cleanup Goals for Soils at Residences Near the Celotex Property, Chicago, IL, 26 January 1996.

Alceon, 1996, PRA
Alceon Corporation, 1996, Probabilistic Calculations to Estimate Risk-Based Cleanup Goals for Soils at Residences Near the Celotex Property, Chicago, IL, 23 February 1996.

Anderson et al, 1984
Anderson, E., N. Browne, S. Duletsky, and T. Warn, 1984, Development of Statistical Distributions or Ranges of Standard Factors Used in Exposure Assessments, US EPA 68-02-3510.

ATSDR, 1990
US Agency for Toxic Substances and Disease Registry, 1990, Toxicological Profile for Polycyclic Aromatic Hydrocarbons, US Public Health Service, Athens, GA.

Benjamin & Cornell, 1970
Benjamin, J.R. and C.A. Cornell, 1970, Probability, Statistics, and Decisions for Civil Engineers, McGraw Hill, New York, NY

Binder et al, 1986
Binder, S., D. Sokal, and D. Maughan, 1986, Estimating Soil Ingestion: The Use of Tracer Elements in Estimating the Amount of Soil Ingested by Young Children, Archives of Environmental Health, Vol. 41, pp 341 - 345.

Bonaccorsi et al, 1984
Bonaccorsi, A.A., et al., 1984, The Influence of Soil Particle Adsorption on 2,3,7,8-TCDD on Biological Uptake in Rabbit, Archives of Toxicology Supplement 7, pp 431 - 434.

Bradley et al, 1994
Bradley, L.J.N. B.H. Magee, and S.L. Allen, 1994, Background Levels of Polycyclic Aromatic Hydrocarbons (PAH) and Selected Metals in New England Urban Soils, Journal of Soil Contamination, Volume 3, Number 4, pp 349 - 361.

Brainard & Burmaster, 1992
Brainard, J. and D.E. Burmaster, 1992, Bivariate Distributions for Height and Weight of Men and Women in the United States, Risk Analysis, 1992, Volume 12, Number 2, pp 267 - 275

Browner, 1995
Browner, C. 1995, Policy Statement on Risk Assessment, Administrator, US Environmental Protection Agency, 21 March 1995.

Brune et al, 1981
Brune, H., R.P. Deutsch-Wenzel, M. Habs, S. Ivankovic, and D. Schmhl, 1981, Investigation of the Tumorigenic Response to Benzo(a)pyrene in Aqueous Caffeine Solution Applied Orally to Sprague-Dawley Rats, J. Cancer Res. Clin. Oncol. 102(2):153-157.

Burmaster & Harris, 1993
Burmaster, D.E. and R.H. Harris, 1993, The Magnitude of Compounding Conservatisms in Superfund Risk Assessments, Risk Analysis, Volume 13, Number 2, pp 131 - 143.

Burmaster, Lloyd & Crouch, 1994
Burmaster, D.E., K.J. Lloyd, and E.A.C. Crouch, 1994, LogNormal Distributions of Body Weight as a Function of Age for Female and Male Children in the United States, Risk Analysis, in revision

Burmaster, Lloyd & Thompson, 1995
Burmaster, D.E., K.J. Lloyd, and K.M. Thompson, 1995, The Need for New Methods to BackCalculate Soil CleanUp Targets in Interval and Probabilistic Risk Assessments, Human and Ecological Risk Assessment, Volume 1, Number 1, pp 89 - 100

Burmaster & Thompson, 1995
Burmaster, D.E. and K.M. Thompson, 1995, BackCalculating CleanUp Targets in Probabilistic Risk Assessments When the Acceptability of Cancer Risk is Defined Under Different Risk Management Policies, Human and Ecological Risk Assessment, Volume 1, Number 1, pp 101 - 120

Burmaster & Anderson, 1994
Burmaster, D.E. and P.D. Anderson, 1994, Principles of Good Practice for the Use of Monte Carlo Techniques in Human Health and Ecological Risk Assessments, Risk Analysis, Volume 14, Number 4, pp 477 - 481

Burmaster & von Stackelberg, 1991
Burmaster, D.E. and K. von Stackelberg, 1991, Using Monte Carlo Simulations in Public Health Risk Assessment: Estimating and Presenting Full Distributions of Risk, Journal of Exposure Analysis and Environmental Epidemiology, Volume 1, Number 4, pp 491 - 512

Burmaster & Bloomfield, 1996
Burmaster, D.E. and L.R. Bloomfield, 1996, Mathematical Properties of the Risk Equation When Variability is Present, Human and Ecological Risk Assessment, Volume 2, Number 2, pp 348 - 355.

Burmaster, 1996
Burmaster, D.E., 1996, Using Beta Distributions Efficiently in a Probabilistic Exposure Assessment, Risk Analysis, in review

Calabrese et al, 1989
Calabrese, E.J. et al, 1989, How Much Soil Do Young Children Ingest: An Epidemiologic Study, In: Petroleum Contaminated Soils, Lewis Publishers, Chelsea, MI, pp 363 - 397.

Clemen, 1991
Clemen, R.T., 1991, Making Hard Decisions, Duxbury Press, Wadsworth Publishing Company, Belmont, CA

Costeff, 1966
Costeff, H., 1966, A Simple Empirical Formula for Calculating Approximate Surface Area in Children, Arch. Dis. Childh., Volume 41, pp 681 - 683.

Crow & Shimizu, 1988
Crow, E.L. and K. Shimizu, Eds., 1988, Lognormal Distributions, Theory and Applications, Marcel Dekker, New York, NY

Decisioneering, 1992
Decisioneering, Inc. 1992. Users Manual for Crystal Ball, Boulder, CO

Draper & Smith, 1981
Draper, N.R. and H. Smith, 1981, Applied Regression Analysis, Second Edition, John Wiley & Sons, New York, NY

E&E, 1995, Report
Ecology & Environment, 1995, Streamlined Human Health Risk Evaluation Celotex Site, Chicago, IL, October 1995.

E&E, 1995, Letter
Ecology & Environment, 1995, Letter dated 5 December 1995 to Richard Boice, US EPA Region V from Dean Tiebout.

ERM, 1995, DSR
Environmental Resources Management, 1995, Data Report Residential Soil Sampling, Celotex Site, Chicago, IL.

ERM, 1995, SSP
Environmental Resources Management, 1995, Support Sampling Plan, Celotex Site, Chicago, IL.

Evans, Hastings & Peacock, 1993
Evans, M., N. Hastings, and B. Peacock, 1993, Statistical Distributions, Second Edition, John Wiley & Sons, New York, NY

Finkel & Golding, 1994
Finkel, A.M. and D. Golding, eds, 1994, Worst Things First, The Debate over Risk-Based National Environmental Priorities, Resources for the Future, Washington, DC.

Fishman, 1996
Fishman, G.S., 1996, Monte Carlo: Concepts, Algorithms, and Applications, Springer-Verlag, New York, NY

Gilbert, 1987
Gilbert, R.O., 1987, Statistical Methods for Environmental Pollution Monitoring, Van Nostrand Reinhold, New York, NY

Goon et al, 1990
Goon, D. N.S. Hatoum, J.D. Jernigan, S.L. Schmitt and P.J. Garvin, 1990, Pharmacokinetics and Oral Bioavailability of Soil-Adsorbed Benzo(a)pyrene (B(a)P) in rats, Abstract. Toxicologist, 10:872.

Goon et al, 1991
Goon, D., N.S. Hatoum, M.J. Klan, J.D. Jernigan, and R.G. Farmer, 1991, Oral Bioavailability of "Aged" Soil-Adsorbed Benzo[a]pyrene (B(a)P) in Rats, Abstract. Toxicologist 11:1356.

Green & Silverman, 1994
Green, P.J. and B.W. Silverman, 1994, Nonparametric Regression and Generalized Linear Models, A Roughness Penalty Approach, Chapman & Hall, London, UK

GRI, 1995
Gas Research Institute, 1995, Draft Report Environmentally Acceptable Endpoints in Soil: Risk-Based Approach to Contaminated Site Management Based on Availability of Chemicals in Soil, Chicago, IL.

Harris & Burmaster, 1992
Harris, R.H. and D.E. Burmaster, 1992, Restoring Science to Superfund Risk Assessment, Toxics Law Reporter, Bureau of National Affairs, Washington, DC, 25 March 1992.

Hatzinger & Alexander, 1995,
Hatzinger, P.B. and M. Alexander, 1995, Effect of Aging of Chemicals in Soil on Their Biodegradability and Extractability, Environmental Science & Technology, Vol 29, pp 537 - 545.

Hawley, 1985
Hawley, J.K., 1985, Assessment of Health Risk from Exposure to Contaminated Soil, Risk Analysis, Vol. 5, pp 289 - 302.

Hecht et al, 1979
Hecht, S.S., W. Grabowski, and K. Groth, 1979, Analysis of Feces for Benzo(a)pyrene After Consumption of Charcoal-Broiled Beef by Rats and Humans, Cosmet. Toxicol., 17:223-227.

Israel & Nelson, 1992
Israeli, M. and C.B. Nelson, 1992, Distributions and Expected Time of Residence for US Households, Risk Analysis, Volume 12, Number 1, pp 65 - 72.

Kissel et al, 1996
Kissel, J.C., K.Y. Richter, and R.A. Fenske, 1996, Field Measurement of Dermal Soil Loading Attributable to Various Activities: Implications for Exposure Assessment, Risk Analysis, Volume 16, Number 1, pp 115 - 125.

LaGoy, 1987
LaGoy, P.K., 1987, Estimated Soil Ingestion Rates for Use in Risk Assessment, Risk Analysis, Vol 7, No. 3, pp 355-359.

Mackay et al., 1992
Mackay, D., Shiu, W. Y., Ma, K.C., 1992, Illustrated Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals, Vol. I and II, Lewis Publishers, Boca Raton, FL.

MA DEP, 1993
Massachusetts Department of Environmental Protection, 1993, Massachusetts Contingency Plan, 310 CMR 40.0000, Bureau of Waste Site Cleanup, 19 November 1993

Mood et al, 1974
Mood, A.M., F.A. Graybill, and D.C. Boes, 1974, Introduction to the Theory of Statistics, Third Edition, McGraw Hill, New York, NY

Morgan, 1984
Morgan, J.T.M., 1984, Elements of Simulation, Chapman and Hall, London, UK

Morgan & Henrion, 1990
Morgan, M.G. and M. Henrion, 1990, Uncertainty, Cambridge University Press, Cambridge, UK

Murray & Burmaster, 1992
Murray, D.M. and D.E. Burmaster, 1992, Estimated Distributions for Total Body Surface Area of Men and Women in the United States, Journal of Exposure Analysis and Environmental Epidemiology, Volume 2, Number 4, pp 451 - 461.

NAS, 1994
National Academy of Sciences, 1994, Science and Judgment in Risk Assessment, National Academy Press, Washington, DC

NCRP, 1996
National Council on Radiation Protection and Measurement, 1996, A Guide for Uncertainty Analysis in Dose and Risk Assessments Related to Environmental Contamination, NCRP Commentary, Number 14, Washington, DC, 10 May 1996

Neal & Rigdon, 1967
Neal, J. and R.H. Rigdon, 1967, Gastric Tumors in Mice Fed Benzo(a)pyrene: A Quantitative Study, Texas Rep. Biol. Med. 25:553.

Ogden, 1996
Ogden Environmental, 1996, Absorption Adjustment Factors for Polycyclic Aromatic Hydrocarbons (PAHs), Westford, MA.

Ott, 1995
Ott, W.R., 1995, Environmental Statistics and Data Analysis, Lewis Publishers, Boca Raton, FL

Ott, 1990
Ott, W.R., 1990, A Physical Explanation of the Lognormality of Pollutant Concentrations, Journal of the Air and Waste Management Association, Volume 40, pp 1378 et seq.

Parzen, 1960
Parzen, E., 1960, Modern Probability Theory and Its Applications, John Wiley & Sons, New York, NY.

Podowski, 1996
Podowski, A., Ph.D., Staff Risk Assessor, personal communication during a meeting at Region V.

Rubinstein, 1981
Rubinstein, R.Y., 1981, Simulation and the Monte Carlo Method, John Wiley & Sons, New York, NY.

Smith et al, 1992
Smith, A.E., P.B. Ryan, and J.S. Evans, 1992, The Effect of Neglecting Correlations When Propagating Uncertainty and Estimating Population Distribution of Risk, Risk Analysis, Volume 12, Number 4, pp 467 - 474, December 1992

Tarter & Lock, 1993
Tarter, M.E. and M.D. Lock, 1993, Model-Free Curve Estimation, Chapman & Hall, New York, NY

Thompson et al, 1992
Thompson, K.M., D.E. Burmaster, and E.A.C. Crouch, 1992, Monte Carlo Techniques for Quantitative Uncertainty Analysis in Public Health Risk Assessments, Risk Analysis, Volume 12, Number 1, pp 53 - 63

Thompson & Burmaster, 1991
Thompson, K.M. and D.E. Burmaster, 1991, Parametric Distributions for Soil Ingestion by Children, Risk Analysis, Volume 11, Number 2, pp 339 - 342.

Travis & Bowers, 1990
Travis, C. and D. Bowers, 1990, Altered Pharmacokinetics of Soil-Adsorbed Benzene Administered Orally to the Rat, Bulletin of Environmental Contamination Toxicology, Vol 45, pp 197 - 207.

US BOC, 1991
US Bureau of the Census, 1991, *Statistical Abstract of the United States: 1991*, 111th edition, Washington, DC.

US DOC, 1992
US Department of Commerce, 1992, Daily Normals of Temperature, Precipitation and Heating and Cooling Degree Days, 1961 - 1990.

US EPA, 1986, FR
US Environmental Protection Agency, 1986, 51 Federal Register 33999 et seq., Wednesday, 24 September 1986.

US EPA, 1986, Mixtures
US Environmental Protection Agency, 1986 Guidelines for the Health Risk Assessment of Chemical Mixtures, 51 FR 34013-34025, 24 September 1986.

US EPA, 1988
US Environmental Protection Agency, 1988, Public Health Risk Evaluation Database, Office of Solid Waste and Emergency Response.

US EPA, 1989, HHEM
US Environmental Protection Agency, Risk Assessment Guidance for Superfund, Volume I, Human Health Evaluation Manual (Part A), Interim Final, US EPA/540/1-89-002, December 1989.

US EPA , 1990, EFH
US Environmental Protection Agency, 1989, Final Report Exposure Factors Handbook, Office of Health and Environmental Assessment, US EPA/600/8-89/043, March 1990.

US EPA, 1990, NCP
US Environmental Protection Agency, 1990, National Oil and Hazardous Substances Pollution Contingency Plan, revisions to 40 CFR Part 300, published as 55 FR 866 et seq., 8 March, 1990.

US EPA, 1991, Default
US Environmental Protection Agency, 1991, Memorandum on Interim Final Standard Default Exposure Factors, Office of Solid Waste and Emergency Response, Washington, DC, 25 March 1991.

US EPA, 1992, RC
US Environmental Protection Agency 1992, Guidance for Risk Characterization for Risk Managers and Risk Assessors, Memorandum signed by F. Henry Habicht, Deputy Administrator, 26 February 1992

US EPA, 1992, EPC
US Environmental Protection Agency, 1992, Supplemental Guidance to RAGS: Calculating the Concentration Term, Publication 9285.7-08, Office of Solid Waste, Washington, DC

US EPA, 1992, Dermal
US Environmental Protection Agency, 1992, Dermal Exposure Assessment: Principles and Applications, US EPA/600/8-91/011B, January 1992.

US EPA, 1992, Exposure
US Environmental Protection Agency, 1992, Guidelines for Exposure Assessment, 57 Federal Register, pp 22888 et seq., 29 May 1992.

US EPA, 1992, SAB
US Environmental Protection Agency, Science Advisory Board's Review of the Draft Final Exposure Assessment Guidelines, EAP-SAB-IAQC-92-015,Science Advisory Board, Washington, DC, 13 January 1992.

US EPA, 1993, PAHs
US Environmental Protection Agency, 1993, Provisional Guidance for Quantitative Risk Assessment of Polycyclic Aromatic Hydrocarbons, US EPA/600/R-93/089.

US EPA, 1993, DW
US Environmental Protection Agency, Office of Water, Drinking Water Regulations and Health Advisories 822/F-93-011, December, 1993.

US EPA ,1994, MC
US Environmental Protection Agency, 1994 Use of Monte Carlo Simulations in Risk Assessments. Region III Technical Guidance Manual, US EPA, Region III, Philadelphia, PA. EPA 903-F-94-001

US EPA, 1994, HEAST
US Environmental Protection Agency, Health Effects Assessment Summary Tables, Annual FY 1994, EPA540/R-94.020, March 1994.

US EPA, 1995, IRIS
US Environmental Protection Agency, US EPA's Database documented in: Integrated Risk Information System, Volume 1 and Electronic Information System, Office of Health and Environmental Assessment, US EPA/600/8-86/032a, March 1987; Downloaded November 1995.

US EPA, 1995, EFH2
US Environmental Protection Agency, 1995, Review Draft Exposure Factors Handbook, US EPA/600/P-95/002A, Washington, DC.

US EPA, 1995, RPR
US Environmental Protection Agency, Risk Policy Report, Vol. 2, No. 1, 20 January 1995.

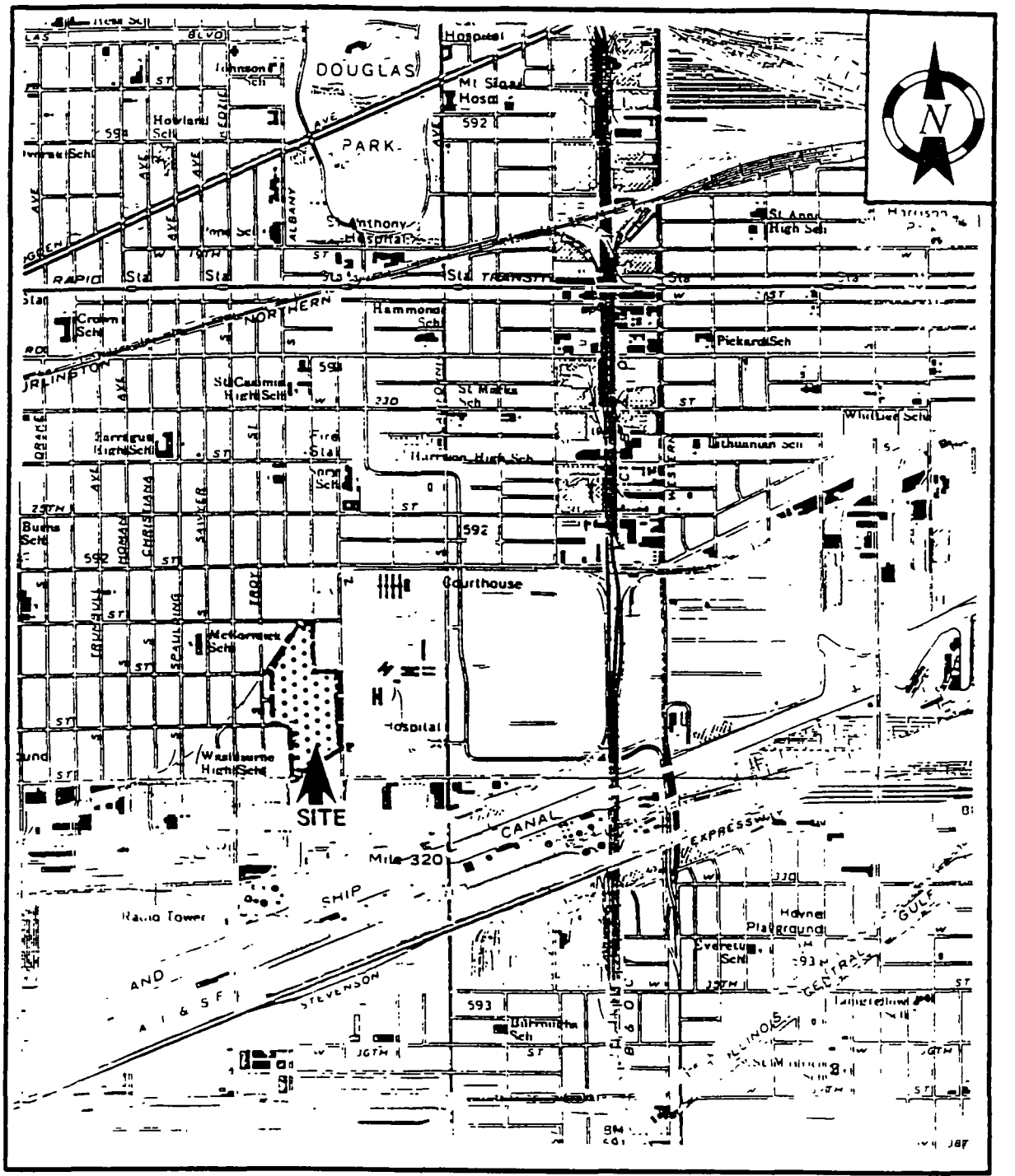
US EPA, 1995, TG
US Environmental Protection Agency, Use of Monte Carlo Simulation in Risk Assessments, Region 8 Superfund Technical Guidance, Number RA-10, , Region VIII, Denver, CO, September 1995

US EPA, 1996, CG
US Environmental Protection Agency, Proposed Guidelines for Carcinogenic Risk Assessment, Office of Research and Development EPA/600/P-92/003C, April 1996, Washington, DC

van Wijnen, 1990
van Wijnen, J.H. et al, 1990, Estimated Soil Ingestion by Children, Environmental Research, Vol 51, pp 147 - 162

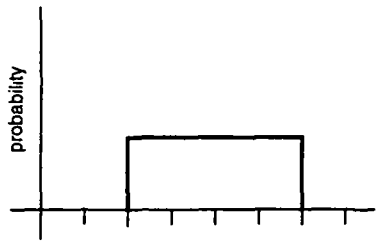
Verschueren, 1983
Verschueren, K., 1983, Handbook of Environmental Data on Organic Chemicals, Second Edition, Van Nostrand Reinhold Company, New York, NY.

NB See also references in each Appendix

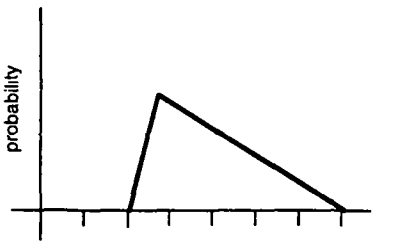



SOURCE USGS Englewood IL Quadrangle, 7 5 Minute Series, 1963, Photorevised 1972 and 1980 SCALE 1 24000

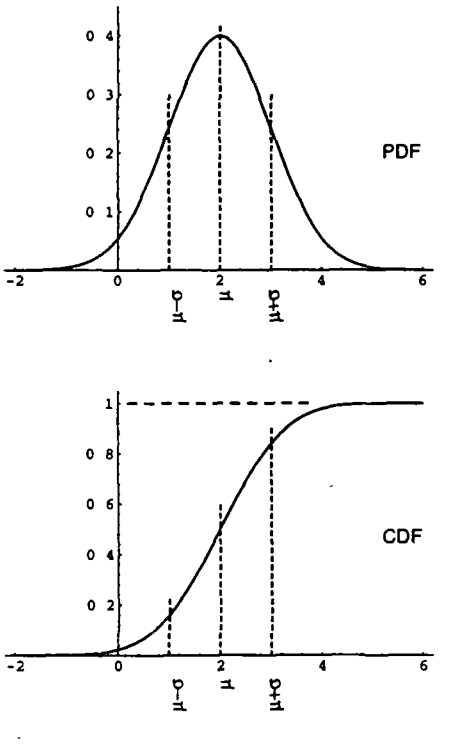
Figure 2-1
SITE LOCATION



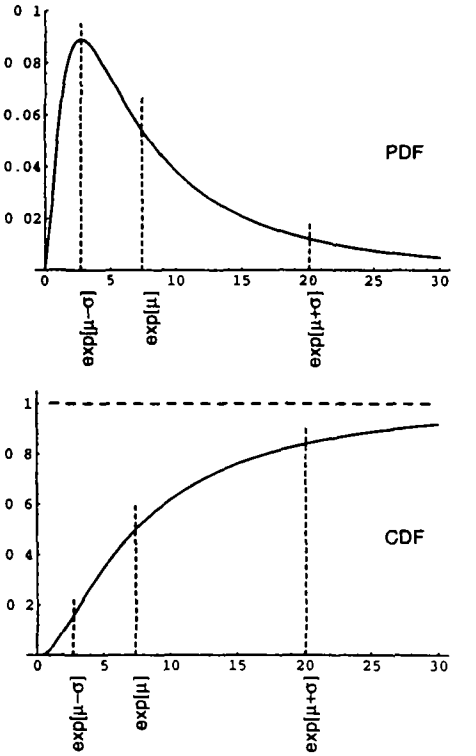

PDF for the Uniform Distribution
U(min, max) = U(2, 6)




PDF for the Triangular Distribution
T(min, mode, max) = T(2, 3, 7)

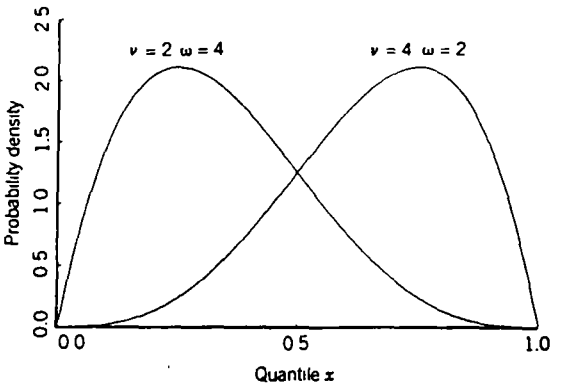



PDF and CDF for the Normal Distribution
N(μ, σ) = N(2, 1)

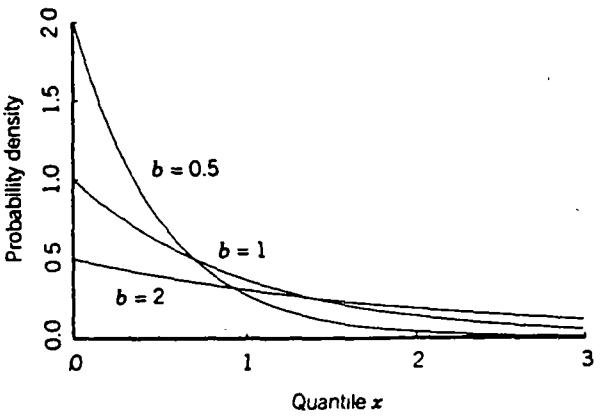



PDF and CDF for the LogNormal Distribution
exp[N(μ, σ)] = exp[N(2, 1)]

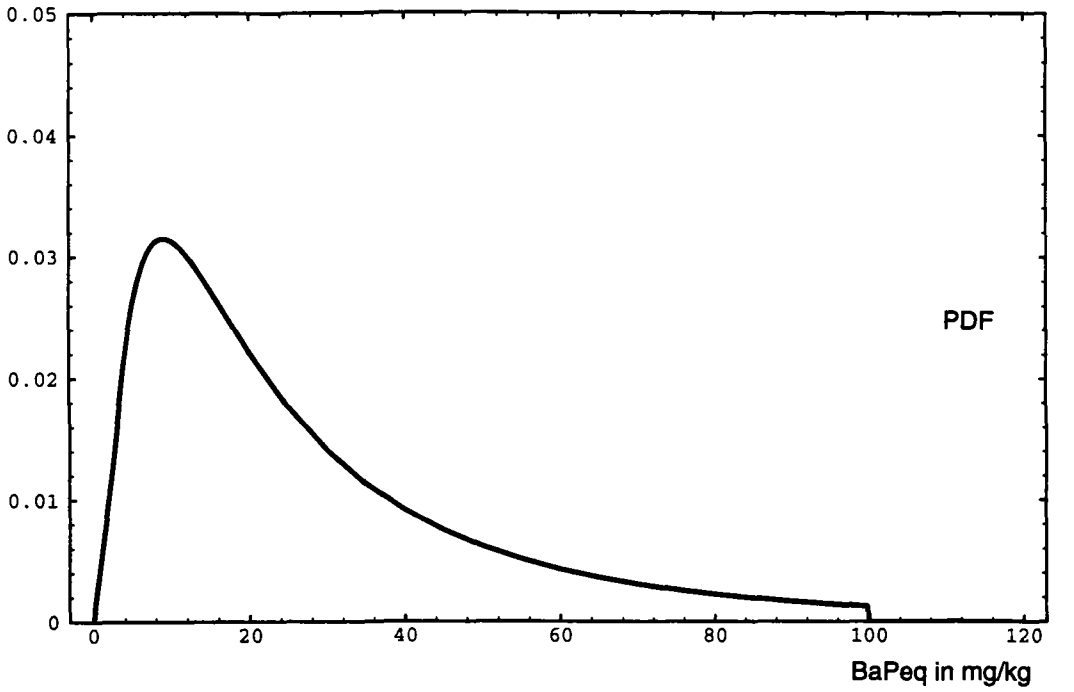
Figure 1
Commonly Used Probability Distributions




*Probability density function for the beta variate* $\beta$: $\nu$, $\omega$.




*Probability density function for the exponential variate* $E$ $b$



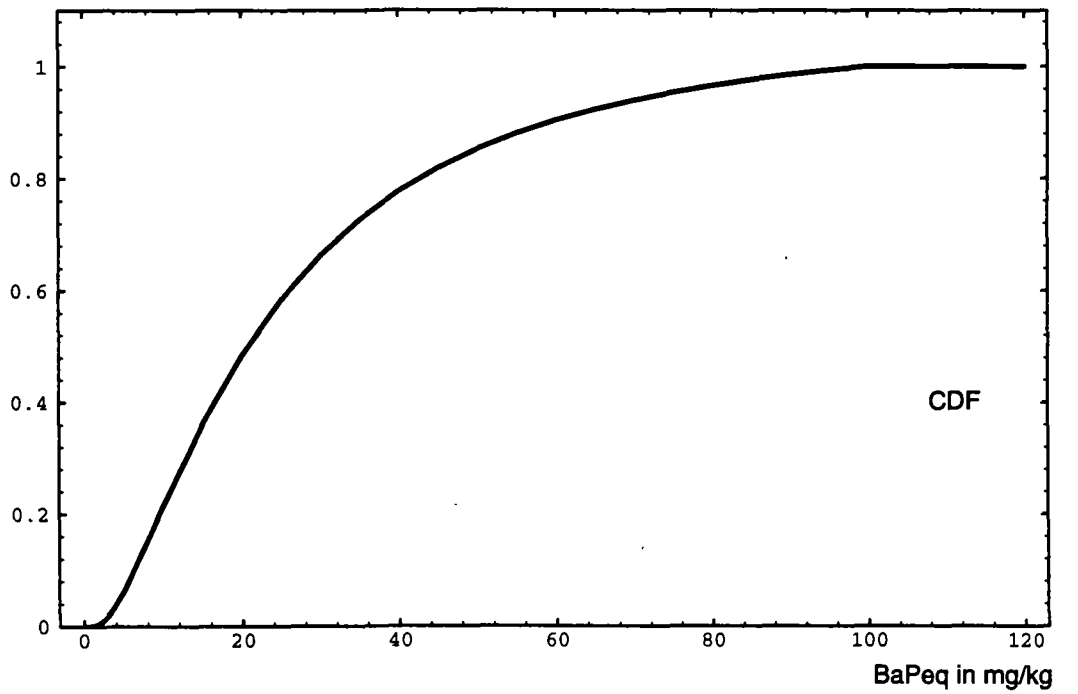




Figure 2: PDF and CDF for the Soil Cleanup Targets for BaPeq (mg/kg) in Surface Soils. See Text for Discussion

Table 1
Study Chemicals, Acronyms, and Synonyms

| | Study Chemical | Synonym | CAS Number |
|---|---|---|---|
| cPAHs | benz(a)anthracene | | 56-55-3 |
| | benzo(a)pyrene | | 50-32-8 |
| | benzo(b)fluoranthene | | 205-99-2 |
| | benzo(k)fluoranthene | | 207-08-9 |
| | chrysene | 1,2-benzophenanthrene | 218-01-9 |
| | dibenz(a,h)anthracene | | 53-70-3 |
| | indeno(1,2,3-cd)pyrene | 2,3-phenylenepyrene | 193-39-5 |

Table 2
Physical and Chemical Properties of the Study Chemicals

| Study Chemical | Molecular Weight (g/mole) | Water Solubility (mg/l) | Vapor Pressure (mm Hg) | Henry's Law Constant (atm•m3/mol) | Koc (-) | Log 10 Kow (-) |
|---|---|---|---|---|---|---|
| benz(a)anthracene | 228 | 5.70E-03 | 2.20E-08 | 1.16E-06 | 1.38E+06 | 5.60 |
| benzo(a)pyrene | 252 | 1.20E-03 | 5.60E-09 | 1.55E-06 | 5.50E+06 | 6.06 |
| benzo(b)fluoranthene | 252 | 1.40E-02 | 5.00E-07 | 1.19E-05 | 5.50E+05 | 6.06 |
| benzo(k)fluoranthene | 252 | 4.30E-03 | 5.10E-07 | 3.94E-05 | 5.50E+05 | 6.06 |
| chrysene | 228 | 1.80E-03 | 6.30E-09 | 1.05E-06 | 2.00E+05 | 5.61 |
| dibenz(a,h)anthracene | 278 | 5.00E-04 | 1.00E-10 | 7.33E-08 | 3.30E+06 | 6.80 |
| indeno(1,2,3-cd)pyrene | 276 | 5.30E-04 | 1.00E-10 | 6.86E-08 | 1.60E+06 | 6.50 |

Note:

All information from US EPA, 1988, Public Health Risk Evaluation Database, Office of Solid Waste & Emergency Response

Table 3
Summary of Key Toxicologic Properties of the Study Chemicals

| Study Chemical | Carcinogenic Toxicity by Ingestion: Drinking Water Unit Risk (µg/l)-1 | Toxic Equivalency Factor (compared to B(a)P) | Cancer Slope Factor (mg / (kg•d))-1 | US EPA Weight of Evidence | Species Tested | Cancer Type | Ref | note |
|---|---|---|---|---|---|---|---|---|
| benzo(a)anthracene | | 1.00E-01 | 7.30E-01 | B2 | mouse | tumors by multiple routes | a | [1] |
| benzo(a)pyrene | 2.10E-04 | 1.00E+00 | 7.30E+00 | B2 | many | tumors by multiple routes | a | [1] |
| benzo(b)fluoranthene | | 1.00E-01 | 7.30E-01 | B2 | mouse | tumors by multiple routes | a | [1] |
| benzo(k)fluoranthene | | 1.00E-02 | 7.30E-02 | B2 | mouse | lung implantation tumors | a | [1] |
| chrysene | | 1.00E-03 | 7.30E-03 | B2 | mouse | carcinomas, lymphomas | a | [1] |
| indeno(1,2,3-c,d)pyrene | | 1.00E-01 | 7.30E-01 | B2 | mouse | tumors from lung implants | a | [1] |

Sources:
a US EPA, 1994, IRIS (May,1994-January,1995)

Notes:
[1] CSFs for non-B(a)P cPAH compunds were derived using the following equation: TEF (cPAH) • CSF (B(a)P) =CSF (cPAH); US EPA, 1993, PAH.

Table 3, continued
Summary of Key Toxicologic Properties of the Study Chemicals

| | Carcinogenic Toxicity by Inhalation | | | | | | |
|---|---|---|---|---|---|---|---|
| Study Chemical | Inhalation Unit Risk (µg/m3)-1 | Cancer Slope Factor (mg / (kg•d))-1 | US EPA Weight of Evidence | Species Tested | Cancer Type | Ref | note |
| benzo(a)anthracene | | 7.30E-01 | B2 | | | a | [1] |
| benzo(a)pyrene | | 7.30E+00 | B2 | | | a | [1] |
| benzo(b)fluoranthene | | 7.30E-01 | B2 | | | a | [1] |
| benzo(k)fluoranthene | | 7.30E-02 | B2 | | | a | [1] |
| chrysene | | 7.30E-03 | B2 | | | a | [1] |
| indeno(1,2,3-c,d)pyrene | | 7.30E-01 | B2 | | | a | [1] |

Table 4
US EPA's Weight-of-Evidence Classification for Carcinogens

| Human Evidence | Animal Evidence: Sufficient | Limited | Inadequate | No Data | No Evidence |
|---|---|---|---|---|---|
| Sufficient | A | A | A | A | A |
| Limited | B1 | B1 | B1 | B1 | B1 |
| Inadequate | B2 | C | D | D | D |
| No Data | B2 | C | D | D | E |
| No Evidence | B2 | C | D | D | E |

Group A = Human Carcinogen (sufficient evidence from epidemiologic studies)

Group B1 = Probable Human Carcinogen (limited evidence from epidemiologic studies)

Group B2 = Probable Human Carcinogen (sufficient animal evidence in absence of adequate human data)

Group C = Possible Human Carcinogen (limited animal evidence in absence of adequate human data)

Group D = Not Classifiable as to Human Carcinogenicity

Group E = Evidence of Non-Carcinogenicity for Humans

Source: US EPA, 51 FR 34000, September 24, 1986
Note: This classification scheme is currently under revision by the US EPA.

Table 5
Estimated Order of Potential Potencies of Selected PAHs
Based on Mouse Skin Carcinogenesis

| Compound | Relative Potency | Reference |
|---|---|---|
| benzo(a)pyrene | 1.0 | |
| benz(a)anthracene | 0.1 | Bingham and Falk, 1969 |
| benzo(b)fluoroanthene | 0.1 | Habs et al., 1980 |
| benzo(k)fluoroanthene | 0.01 | Habs et al., 1980 |
| chrysene | 0.001 | Wynder and Hoffmann, 1959 |
| dibenz(a,h)anthracene | 1.0 | Wynder and Hoffmann, 1959 |
| indeno(1,2,3-cd)pyrene | 0.1 | Habs et al., 1980; Hoffmann and Wynder, 1966 |

Source:
Provisional Guidance for Quantitative Risk Assessment
of Polycyclic Aromatic Hydrocarbons
US EPA, 1993, EPA/600/R-93/089

Table 6
Summary of Exposure Scenarios to Estimate Health Risks

| Variable | Conditions<br>Type of receptor: | Residents in Vicinity of Celotex Property<br>current / future<br>child | <br>current / future<br>teenager | <br>current / future<br>adult |
|---|---|---|---|---|
| | Age during exposure (yr): | 1 through 6 | 7 through 17 | ≥ 18 |
| | Average lifetime (yr): | 70 | 70 | 70 |
| Ingestion | | | | |
| | Incidental Ingestion of soil | Q | Q | Q |
| Inhalation | | | | |
| | Inhalation of fugitive dust | q | q | q |
| | Inhalation of soil vapors | NE | NE | NE |
| Dermal | | | | |
| | Dermal contact with soils | Q | Q | Q |

Notes:
Q = Exposure pathway quantified
q = Exposure pathway evaluated qualitatively
NE = Exposure pathway not evaluated

Table 7
Detailed Exposure Scenarios to Estimate Health Risks

| | | | Celotex's + AlliedSignal's Point Values | | | Distribution | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Oral Ingestion** | | **Units** | **Point Estimate** | **RME (HEE)** | **Source** | **Distribution** | **Support** | **95th Percentile** | **Median** |
| BaPeq - outdoor soil | SoilBaPeq | mg/kg | | | A | | | | |
| BaPeq - indoor dust | DustBaPeq | mg/kg | | | B | | | | |
| Transfer Coefficient | TC | - | | 0 42 | C | exp[N(-0 877,0 366)] | (0, 1) | 0 74 | 0 42 |
| Absorption Adjustment Factor - oral | AAFo | - | | 0 27 | D | B4[1,3,0 945,0 07] | (0, 1) | 0 66 | 0 27 |
| Soil Ingestion Rate - child | SIRc | mg/d | | 200 | E | exp[N(4 13,0 8)] | (0, ∞) | 232 33 | 62 18 |
| Soil Ingestion Rate - teen | SIRt | mg/d | | 100 | E | exp[N(3 44,0 8)] | (0, ∞) | 116 48 | 31 22 |
| Soil Ingestion Rate - adult | SIRa | mg/d | | 100 | E | exp[N(3 44,0 8)] | (0, ∞) | 116 35 | 31 16 |
| Conversion factor (mg->kg) | CF | kg/mg | 1 00E-06 | | | Constant | | | |
| Conversion factor (m2->cm2) | CFs | cm2/m2 | 1 00E+04 | | | Constant | | | |
| Body Weight - child | BWc | kg | | 16 4 | F | exp[N(2 69,0 33)] | (0, ∞) | 25 35 | 14 73 |
| Body Weight - teen | BWt | kg | | 44 9 | F | exp[N(3 75,0 37)] | (0, ∞) | 78 18 | 42 50 |
| Body Weight - adult | BWa | kg | | 70 | G | exp[N(4 263,0.206)] | (0, ∞) | 99 69 | 71 03 |
| Exposure Frequency - outdoor - child | EFOc | d | | 164 | H | Custom | | 246 | 164 |
| Exposure Frequency - indoor - child | EFIc | d | | 186 | H | [simulated] | | 254 | 186 |
| Exposure Frequency - outdoor - teen | EFOc | d | | 177 | H | Custom | | 264 | 177 |
| Exposure Frequency - indoor - teen | EFIc | d | | 173 | H | [simulated] | | 245 | 173 |
| Exposure Frequency - outdoor - adult | EFOc | d | | 167 | H | Custom | | 246 | 167 |
| Exposure Frequency - indoor - adult | EFIc | d | | 183 | H | [simulated] | | 255 | 183 |
| Days per Year | dpy | d/yr | 350 | | I | Point Estimate | | | |
| Exposure Duration - child | EDc | yr | | 6 | J | [simulated] | | 6 | 3 2 |
| Exposure Duration - teen | EDt | yr | | 11 | J | [simulated] | | 11 | 0 2 |
| Exposure Duration - adult | EDa | yr | | 1 | J | [simulated] | | 7 084 | 0 1 |
| Years in Lifetime | Lifetime | yr | 70 | | K | Point Estimate | | | |
| **Dermal Contact** | | **Units** | **Point Estimate** | **RME (HEE)** | **Source** | **Distribution** | **Support** | **95th Percentile** | **Median** |
| BaPeq - outdoor soil | SoilBaPeq | mg/kg | | | A | | | | |
| BaPeq - indoor dust | DustBaPeq | mg/kg | | | B | | | | |
| Transfer Coefficient | TC | - | | 0 42 | C | exp[N(-0 877,0 366)] | (0, 1) | 0 74 | 0 41 |
| Absorption Adjustment Factor - dermal | AAFd | - | | 0 02 | D | *** | (0, ∞) | 0 07 | 0 02 |
| Soil Adherence Rate | SAR | mg/(cm2•d) | | 1 | L | exp[N(-1 71,1 01)] | (0, ∞) | 0 952 | 0 181 |
| Dust Adherence Rate | DAR | mg/(cm2•d) | | 0 2 | M | exp[N(-3 1,1 01)] | (0, ∞) | 0 24 | 0 05 |
| Skin Surface Area - child | SAc | m2 | | 0 73 | N | [simulated] | | 0 941 | 0 629 |
| Skin Surface Area - teen | SAt | m2 | | 1 5 | N | [simulated] | | 1 901 | 1 336 |
| Skin Surface Area - adult | SAa | m2 | | 2 | N | [simulated] | | 2 139 | 1 808 |
| Fraction of Skin Area Exposed | Frc | - | 0 25 | | O | Point Estimate | | | |
| Conversion factor (mg->kg) | CF | kg/mg | 1 00E-06 | | | Constant | | | |
| Conversion factor (m2->cm2) | CFs | cm2/m2 | 1 00E+04 | | | Constant | | | |
| Body Weight - child | BWc | kg | | 16 4 | F | exp[N(2 69,0 33)] | (0, ∞) | 25 35 | 14 73 |
| Body Weight - teen | BWt | kg | | 44 9 | F | exp[N(3 75,0 37)] | (0, ∞) | 78 18 | 42 50 |
| Body Weight - adult | BWa | kg | | 70 | G | exp[N(4 263,0 206)] | (0, ∞) | 99 69 | 71 03 |
| Exposure Frequency - outdoor - child | EFOc | d | | 164 | H | Custom | | 246 | 164 |
| Exposure Frequency - indoor - child | EFIc | d | | 186 | H | [simulated] | | 254 | 186 |
| Exposure Frequency - outdoor - teen | EFOt | d | | 177 | H | Custom | | 264 | 177 |
| Exposure Frequency - indoor - teen | EFIt | d | | 173 | H | [simulated] | | 245 | 173 |
| Exposure Frequency - outdoor - adult | EFOa | d | | 167 | H | Custom | | 246 | 167 |
| Exposure Frequency - indoor - adult | EFIa | d | | 183 | H | [simulated] | | 255 | 183 |
| Days per Year | dpy | d/yr | 350 | | I | Point Estimate | | | |
| Exposure Duration - child | EDc | yr | | 6 | J | [simulated] | | 6 | 3 2 |
| Exposure Duration - teen | EDt | yr | | 11 | J | [simulated] | | 11 | 0 2 |
| Exposure Duration - adult | EDa | yr | | 1 | J | [simulated] | | 7 084 | 0 1 |
| Years in Lifetime | Lifetime | yr | 70 | | K | Point Estimate | | | |

*** B4[1,5,0 147,0] / B4[4,1,0 397,0 603]

Table 7, continued
Detailed Exposure Scenarios to Estimate Health Risks

| | | | US EPA Region V's Point Values | | |
|---|---|---|---|---|---|
| Oral Ingestion | | Units | Point Estimate | default RME | Source |
| BaPeq - outdoor soil | SoilBaPeq | mg/kg | | | AA |
| BaPeq - indoor dust | DustBaPeq | mg/kg | | | AA |
| Transfer Coefficient | TC | - | | 0 42 | AA |
| Absorption Adjustment Factor - oral | AAFo | - | | 0 9 | AA |
| Soil Ingestion Rate - child | SIRc | mg/d | | 200 | AA |
| Soil Ingestion Rate - teen | SIRt | mg/d | | 200 | AA |
| Soil Ingestion Rate - adult | SIRa | mg/d | | 100 | AA |
| Conversion factor (mg->kg) | CF | kg/mg | 1 00E-06 | | AA |
| Conversion factor (m2->cm2) | CFs | cm2/m2 | 1 00E+04 | | AA |
| Body Weight - child | BWc | kg | | 15 | AA |
| Body Weight - teen | BWt | kg | | 45 | AA |
| Body Weight - adult | BWa | kg | | 70 | AA |
| Exposure Frequency - outdoor - child | EFOc | d | | 350 | AA |
| Exposure Frequency - indoor - child | EFIc | d | | 0 | AA |
| Exposure Frequency - outdoor - teen | EFOc | d | | 350 | AA |
| Exposure Frequency - indoor - teen | EFIc | d | | 0 | AA |
| Exposure Frequency - outdoor - adult | EFOc | d | | 350 | AA |
| Exposure Frequency - indoor - adult | EFIc | d | | 0 | AA |
| Days per Year | dpy | d/yr | | 365 | AA |
| Exposure Duration - child | EDc | yr | | 6 | AA |
| Exposure Duration - teen | EDt | yr | | 11 | AA |
| Exposure Duration - adult | EDa | yr | | 13 | AA |
| Years in Lifetime | Lifetime | yr | | 70 | AA |
| Dermal Contact | | Units | Point Estimate | default RME | Source |
| BaPeq - outdoor soil | SoilBaPeq | mg/kg | | | AA |
| BaPeq - indoor dust | DustBaPeq | mg/kg | | | AA |
| Transfer Coefficient | TC | - | | na | AA |
| Absorption Adjustment Factor - dermal | AAFd | - | | 0 15 | AA |
| Soil Adherence Rate | SAR | mg/(cm2-d) | | 1 | AA |
| Dust Adherence Rate | DAR | mg/(cm2-d) | | na | AA |
| Skin Surface Area - child | SAc | m2 | | 0 73 | AA |
| Skin Surface Area - teen | SAt | m2 | | 1 5 | AA |
| Skin Surface Area - adult | SAa | m2 | | 2 | AA |
| Fraction of Skin Area Exposed | Frc | - | 0 25 | | AA |
| Conversion factor (mg->kg) | CF | kg/mg | 1 00E-06 | | AA |
| Conversion factor (m2->cm2) | CFs | cm2/m2 | 1 00E+04 | | AA |
| Body Weight - child | BWc | kg | | 15 | AA |
| Body Weight - teen | BWt | kg | | 45 | AA |
| Body Weight - adult | BWa | kg | | 70 | AA |
| Exposure Frequency - outdoor - child | EFOc | d | | 350 | AA |
| Exposure Frequency - indoor - child | EFIc | d | | 0 | AA |
| Exposure Frequency - outdoor - teen | EFOt | d | | 350 | AA |
| Exposure Frequency - indoor - teen | EFIt | d | | 0 | AA |
| Exposure Frequency - outdoor - adult | EFOa | d | | 350 | AA |
| Exposure Frequency - indoor - adult | EFIa | d | | 0 | AA |
| Days per Year | dpy | d/yr | | 365 | AA |
| Exposure Duration - child | EDc | yr | | 6 | AA |
| Exposure Duration - teen | EDt | yr | | 11 | AA |
| Exposure Duration - adult | EDa | yr | | 13 | AA |
| Years in Lifetime | Lifetime | yr | | 70 | AA |

Table 7, continued
Detailed Exposure Scenarios to Estimate Health Risks
Ecology & Environment's Report

| Source | |
|---|---|
| A | (the goal of the calculations) |
| B | calculated from transfer coefficient |
| C | Trowbridge & Burmaster, 1996 |
| D | Magee et al, 1996 |
| E | Thompson & Burmaster, 1991, UD EPA, 1995, EFH2, LaGoy, |
| F | Burmaster et al, 1994, Burmaster & Crouch, 1996 |
| G | Brainard & Burmaster, 1992, US EPA, 1995, EFH2 |
| H | Alceon -- see Appendix XXX |
| I | US EPA, 1989, HHEM |
| J | Alceon -- see Appendix XXX |
| K | US EPA, 1989, HHEM |
| L | US EPA, 1992, Dermal |
| M | Kissel et al, 1996, US EPA, 1992, Dermal |
| N | Murray & Burmaster, 1992, US EPA, 1995, EFH2 |
| O | Ecology & Environment, October 1995 |
| AA | Ecology & Environment, October 1995 and discussions with Staff Members of US EPA's Region V |

Table 8
Exposure Frequency as a Function of Temperature

| | 32 degF | 40 degF | 50 degF | 60 degF | 70 degF |
|---|---|---|---|---|---|
| N days at or above | 282 | 246 | 196 | 143 | 86 |
| N days below | 83 | 119 | 169 | 222 | 279 |
| child | 0.00 | 0.05 | 0.20 | 0.70 | 1.00 |
| teen | 0.00 | 0.10 | 0.30 | 0.85 | 1.00 |
| adult | 0.00 | 0.05 | 0.25 | 0.70 | 1.00 |
| checksum | 365 | 365 | 365 | 365 | 365 |

Table 9
Formulae Used to Estimate Doses and Risk

**1. Formula for Average Daily Dose for a Day of Exposure - ADD (day) - Ingestion of Soil or Dust**

$$ADD\ (day)\ \left(\frac{mg}{kg \bullet d}\right) = \frac{EPC\ \left(\frac{mg}{kg}\right) \bullet IngR\ \left(\frac{mg}{d}\right) \bullet CF_{km}\ \left(10^{-6}\ \frac{kg}{mg}\right) \bullet oAAF}{BW\ (kg)}$$

where:

| | | |
|---|---|---|
| ADD (day) | = | Average Daily Dose for a Day of Exposure via Ingestion |
| EPC | = | Acceptable Exposure Point Concentration of Study Chemical in Soil or Dust |
| IngR | = | Average Incidental Ingestion Rate |
| $CF_{km}$ | = | Conversion Factor for kg to mg |
| oAAF | = | Oral Absorption Adjustment Factor |
| BW | = | Body Weight |

**2. Formula for Average Daily Dose for a Day of Exposure - ADD (day) - Dermal Contact with Soil or Dust**

$$ADD\ (d)\ \left(\frac{mg}{kg \bullet d}\right) = \frac{EPC\ \left(\frac{mg}{kg}\right) \bullet AdhR\ \left(\frac{mg}{cm^2 \bullet d}\right) \bullet SA\ (m^2) \bullet FracExp \bullet CF_{km}\ \left(10^{-6}\frac{kg}{mg}\right) \bullet CF_{cm}\ \left(10^4\ \frac{cm^2}{m^2}\right) \bullet dAAF}{BW\ (kg)}$$

where:

| | | |
|---|---|---|
| ADD (d) | = | Average Daily Dose for a Day of Exposure via Dermal Contact |
| EPC | = | Average Concentration of Study Chemical in Soil or Dust |
| AdhR | = | Adherence Rate of Soil on Skin |
| SA | = | Surface Area of Body |
| FracExp | = | Fraction of Body Exposed |
| $CF_{km}$ | = | Conversion Factor for kg to mg |
| $CF_{cm}$ | = | Conversion Factor for $cm^2$ to $m^2$ |
| dAAF | = | Dermal Absorption Adjustment Factor |
| BW | = | Body Weight |

### 3. Formula for Average Daily Dose for a Lifetime of Exposure - ADD (life)

$$\text{ADD (life)}\left(\frac{\text{mg}}{\text{kg} \bullet \text{d}}\right) = \text{ADD (day)}\left(\frac{\text{mg}}{\text{kg} \bullet \text{d}}\right) \bullet \text{EF}\left(\frac{\text{d}}{\text{yr}}\right) \bullet \text{CF}_{\text{dy}}\left(\frac{1\ \text{yr}}{365\ \text{d}}\right)$$

where:

| | | |
|---|---|---|
| ADD (life) | = | Average Daily Dose for a Lifetime of Exposure |
| ADD (day) | = | Average Daily Dose for a Day of Exposure |
| EF | = | Frequency of Exposure in Life Stage |
| $CF_{dy}$ | = | Conversion Factor for d to yr |

### 4. Formula for Total Average Daily Dose - ADD (total)

$$\text{ADD (total)}\left(\frac{\text{mg}}{\text{kg} \bullet \text{d}}\right) = \sum\left(\text{ADD (life)}\left(\frac{\text{mg}}{\text{kg} \bullet \text{d}}\right) \bullet \text{ED(yr)}\right) \bullet \left(\frac{1}{\text{Lifetime}}\right)$$

where:

| | | |
|---|---|---|
| ADD (total) | = | Total Average Daily Dose |
| ADD (life) | = | Average Daily Dose for a Lifetime of Exposure |
| ED | = | Duration of Exposure in Life Stage |
| $CF_{dy}$ | = | Conversion Factor for d to yr |
| Lifetime | = | Years in Lifetime |

Table 10

| Variable | Child age: 1 throu | | Teen age: 7 throu | | Adult age: 18 thro | | Units |
|---|---|---|---|---|---|---|---|
| BaPeq - outdoor soil | SoilBaPeq | =SoilBaPeq | SoilBaPeq | =SoilBaPeq | SoilBaPeq | 27 5 | mg/kg |
| Transfer Coefficient | TC | =TC | TC | =TC | TC | 0 42 | dimensionless |
| BaPeq - indoor dust | DustBaPeq | =DustBaPeq | DustBaPeq | =DustBaPeq | DustBaPeq | =TC*SoilBaPeq | |
| Soil Ingestion Rate | SIRc | =IF(M13=0,O13,P13) | SIRt | =IF(M14=0,O14,P14) | SIRa | =IF(M15=0,O15,P15) | mg/day |
| Absorption Adjustment Factor - oral | AAFo | =AAFo | AAFo | =AAFo | AAFo | =IF(M12=0,O12,P12) | dimensionless |
| Average Daily Dose (day) - oral -outdoor | OADDOc | =SoilBaPeq*AAFo*SIRc*CF/BWc | OADDOt | =SoilBaPeq*AAFo*SIRt*CF/BWt | OADDOa | =SoilBaPeq*AAFo*SIRa*CF/BWa | mg/(kg*day) |
| Average Daily Dose (day) - oral - indoor | OADDIc | =DustBaPeq*AAFo*SIRc*CF/BWc | OADDIt | =DustBaPeq*AAFo*SIRt*CF/BWt | OADDIa | =DustBaPeq*AAFo*SIRa*CF/BWa | mg/(kg*day) |
| Soil Adherence Rate | SAR | =SAR | SAR | =SAR | SAR | =IF(M39=0,O39,P39) | mg/(cm2*d) |
| Dust Adherence Rate | DAR | =DAR | DAR | =DAR | DAR | =IF(M40=0,O40,P40) | mg/(cm2*d) |
| Skin Surface Area | SAc | =IF(M41=0,O41,P41) | SAt | =IF(M42=0,O42,P42) | SAa | =IF(M43=0,O43,P43) | cm2 |
| Fraction of Skin Area Exposed - outdoor | FrcOc | 0 25 | FrcOt | 0 25 | FrcOa | 0 25 | - |
| Fraction of Skin Area Exposed - indoor | Frcic | 0 25 | Frclt | 0 25 | Frcla | 0 25 | - |
| Absorption Adjustment Factor - dermal | AAFd | =AAFd | AAFd | =AAFd | AAFd | =IF(M38=0,O38,P38) | dimensionless |
| Average Daily Dose (day) - dermal - outdoor | DADDOc | =(SoilBaPeq*AAFd*SAR*SAc*FrcOc*CF*CFs)/BWc | DADDOt | =(SoilBaPeq*AAFd*SAR*SAt*FrcOt*CF*CFs)/BWt | DADDOa | =(SoilBaPeq*AAFd*SAR*SAa*FrcOa*CF*CFs)/BWa | mg/(kg*day) |
| Average Daily Dose (day) - dermal - indoor | DADDIc | =(DustBaPeq*AAFd*DAR*SAc*Frcic*CF*CFs)/BWc | DADDIt | =(DustBaPeq*AAFd*DAR*SAt*Frclt*CF*CFs)/BWt | DADDIa | =(DustBaPeq*AAFd*DAR*SAa*Frcla*CF*CFs)/BWa | mg/(kg*day) |
| Body Weight | BWc | =IF(M18=0,O18,P18) | BWt | =IF(M19=0,O19,P19) | BWa | =IF(M20=0,O20,P20) | kg |
| Conversion factor (mg->kg) | CF | =CF | CF | =CF | CF | 0 000001 | kg/mg |
| Conversion factor (m2->cm2) | CFs | =CFs | CFs | =CFs | CFs | 10000 | cm2/m2 |
| Number of Days near Celotex | DayCc | 350 | DayCt | 350 | DayCa | 350 | day/yr |
| Exposure Frequency - outdoor | EFOc | 164 | EFOt | 177 | EFOa | 167 | day/yr |
| Exposure Frequency - indoor | EFIc | =DayCc-EFOc | EFIt | =DayCt-EFOt | EFIa | =DayCa-EFOa | day/yr |
| Exposure Duration | EDc | =IF(M28=0,O28,P28) | EDt | =IF(M29=0,O29,P29) | EDa | =IF(M30=0,O30,P30) | yr |
| Days per Year | dpy | =dpy | dpy | =dpy | dpy | 365 | day/yr |
| Years in Lifetime | Lifetime | =Lifetime | Lifetime | =Lifetime | Lifetime | 70 | yr |
| Averaging time | AT | =AT | AT | =AT | AT | =Lifetime*dpy | day |
| Average Daily Dose(lifetime) - oral | OADDIc | =(OADDOc*EFOc+OADDIc*EFIc)/dpy | OADDIt | =(OADDOt*EFOt+OADDIt*EFIt)/dpy | OADDIa | =(OADDOa*EFOa+OADDIa*EFIa)/dpy | mg/(kg*day) |
| Total Average Daily Dose - oral | | | | | OADD | =(EDc*OADDIc+EDt*OADDIt+EDa*OADDIa)/Lifetime | mg/(kg*day) |
| Average Daily Dose(lifetime) - dermal | DADDIc | =(DADDOc*EFOc+DADDIc*EFIc)/dpy | DADDIt | =(DADDOt*EFOt+DADDIt*EFIt)/dpy | DADDIa | =(DADDOa*EFOa+DADDIa*EFIa)/dpy | mg/(kg*day) |
| Total Average Daily Dose - dermal | | | | | DADD | =(EDc*DADDIc+EDt*DADDIt+EDa*DADDIa)/Lifetime | mg/(kg*day) |
| Cancer Slope Factor - ingestion | CSF | =CSF | CSF | =CSF | CSF | 7 3 | (kg*day)/mg |
| Incremental Lifetime Cancer Risk - oral | | | | | ILCRo | =OADD*CSF | |
| Incremental Lifetime Cancer Risk - dermal | | | | | ILCRd | =DADD*CSF | |
| Incremental Lifetime Cancer Risk - Total | | | | | ILCR | =ILCRo+ILCRd | |

Table 10

| Oral Ingestion | | Toggle | CT Toggle 0 Median | RME Toggle 1 95th percentile | Point Estimate | Units |
|---|---|---|---|---|---|---|
| BaPeq - outdoor soil | SoilBaPeq | | | | | mg/kg |
| BaPeq - indoor dust | DustBaPeq | | | | | mg/kg |
| Transfer Coefficient | TC | | **0.42** | **0.42** | **0.42** | - |
| Absorption Adjustment Factor - oral | AAFo | 1 | **0.27** | **0.27** | | - |
| Soil Ingestion Rate - child | SIRc | 1 | 62 | 200 | | mg/d |
| Soil Ingestion Rate - teen | SIRt | =M13 | 31 | 100 | | mg/d |
| Soil Ingestion Rate - adult | SIRa | =M14 | 31 | 100 | | mg/d |
| Conversion factor (mg->kg) | CF | | | | 0 000001 | kg/mg |
| Conversion factor (m2->cm2) | CFs | | | | 10000 | cm2/m2 |
| Body Weight - child | BWc | 1 | 14 7 | 16 4 | | kg |
| Body Weight - teen | BWt | =M18 | 42 5 | 44 9 | | kg |
| Body Weight - adult | BWa | =M19 | 71 | 70 | | kg |
| Exposure Frequency - outdoor - child | EFOc | 1 | 164 | 164 | | d |
| Exposure Frequency - indoor - child | EFIc | =M21 | 186 | =350-P21 | | d |
| Exposure Frequency - outdoor - teen | EFOc | =M22 | 177 | 177 | | d |
| Exposure Frequency - indoor - teen | EFIc | =M23 | 173 | =350-P23 | | d |
| Exposure Frequency - outdoor - adult | EFOc | =M24 | 167 | 167 | | d |
| Exposure Frequency - indoor - adult | EFIc | =M25 | 183 | =350-P25 | | d |
| Days per Year | dpy | | | | 350 | d/y |
| Exposure Duration - child | EDc | 1 | 3 2 | 6 | | y |
| Exposure Duration - teen | EDt | =M28 | 0 2 | 11 | | y |
| Exposure Duration - adult | EDa | =M29 | 0 1 | 1 | | y |
| Years in Lifetime | Lifetime | | | | 70 | y |

| Dermal Contact | | | Median | 95th percentile | Point Estimate | Units |
|---|---|---|---|---|---|---|
| BaPeq - outdoor soil | SoilBaPeq | | | | | mg/kg |
| BaPeq - indoor dust | DustBaPeq | | | | | mg/kg |
| Transfer Coefficient | TC | | **0.42** | **0.42** | **0.42** | - |
| Absorption Adjustment Factor - dermal | AAFd | =M12 | **0.02** | **0.02** | | - |
| Soil Adherence Rate | SAR | 1 | 0 181 | 1 | | mg/(cm2•d) |
| Dust Adherence Rate | DAR | =M39 | 0 045 | 0 2 | | mg/(cm2•d) |
| Skin Surface Area - child | SAc | 1 | 0 63 | 0 73 | | m2 |
| Skin Surface Area - teen | SAt | =M41 | 1 34 | 1 5 | | m2 |
| Skin Surface Area - adult | SAa | =M42 | 1 8 | 2 | | m2 |
| Fraction of Skin Area Exposed | Frc | | | | 0 25 | - |
| Conversion factor (mg->kg) | CF | | | | 0 000001 | kg/mg |
| Conversion factor (m2->cm2) | CFs | | | | 10000 | cm2/m2 |
| Body Weight - child | BWc | =M18 | 14 7 | 16 4 | | kg |
| Body Weight - teen | BWt | =M19 | 42 5 | 44 9 | | kg |
| Body Weight - adult | BWa | =M20 | 71 | 70 | | kg |
| Exposure Frequency - outdoor - child | EFOc | =M21 | 164 | 164 | | d |
| Exposure Frequency - indoor - child | EFIc | =M22 | 186 | =350-P50 | | d |
| Exposure Frequency - outdoor - teen | EFOt | =M23 | 177 | 177 | | d |
| Exposure Frequency - indoor - teen | EFIt | =M24 | 173 | =350-P52 | | d |
| Exposure Frequency - outdoor - adult | EFOa | =M25 | 167 | 167 | | d |
| Exposure Frequency - indoor - adult | EFIa | =M26 | 183 | =350-P54 | | d |
| Days per Year | dpy | | | | 350 | d/y |
| Exposure Duration - child | EDc | =M28 | 3 2 | 6 | | y |
| Exposure Duration - teen | EDt | =M29 | 0 2 | 11 | | y |
| Exposure Duration - adult | EDa | =M30 | 0 1 | 1 | | y |
| Years in Lifetime | Lifetime | | | | 70 | y |

Table 11
Estimated Deterministic RBCG
RME (HEE) - Celotex + AlliedSignal

| Variable | Child age: 1 through 6 yr | | Teen age: 7 through 17 yr | | Adult age: 18 through 70 yr | | Units |
|---|---|---|---|---|---|---|---|
| BaPeq - outdoor soil | SoilBaPeq | 27 500 | SoilBaPeq | 27 500 | SoilBaPeq | 27 500 | mg/kg BaPeq |
| Transfer Coefficient | TC | 0 420 | TC | 0 420 | TC | 0 420 | dimensionless |
| BaPeq - indoor dust | DustBaPeq | 11 550 | DustBaPeq | 11 550 | DustBaPeq | 11 550 | |
| Soil Ingestion Rate | SIRc | 200 000 | SIRt | 100 000 | SIRa | 100 000 | mg/day |
| Absorption Adjustment Factor - oral | AAFo | 0 270 | AAFo | 0 270 | AAFo | 0 270 | dimensionless |
| Average Daily Dose (day) - oral -outdoor | OADDOc | 9 05E-05 | OADDOt | 1 65E-05 | OADDOa | 1 06E-05 | mg/(kg*day) |
| Average Daily Dose (day) - oral - indoor | OADDIc | 3 80E-05 | OADDIt | 6 95E-06 | OADDIa | 4 46E-06 | mg/(kg*day) |
| Soil Adherence Rate | SAR | 1 000 | SAR | 1 000 | SAR | 1 000 | mg/(cm2*d) |
| Dust Adherence Rate | DAR | 0 200 | DAR | 0 200 | DAR | 0 200 | mg/(cm2*d) |
| Skin Surface Area | SAc | 0 730 | SAt | 1 500 | SAa | 2 000 | cm2 |
| *Fraction of Skin Area Exposed - outdoor* | *FrcOc* | *0 250* | *FrcOt* | *0 250* | *FrcOa* | *0 250* | - |
| Fraction of Skin Area Exposed - indoor | FrcIc | 0 250 | FrcIt | 0 250 | FrcIa | 0 250 | - |
| Absorption Adjustment Factor - dermal | AAFd | 0 020 | AAFd | 0 020 | AAFd | 0 020 | dimensionless |
| Average Daily Dose (day) - dermal - outdoor | DADDOc | 6 12E-05 | DADDOt | 4 59E-05 | DADDOa | 3 93E-05 | mg/(kg*day) |
| Average Daily Dose (day) - dermal - indoor | DADDIc | 5 14E-06 | DADDIt | 3 86E-06 | DADDIa | 3 30E-06 | mg/(kg*day) |
| Body Weight | BWc | 16.400 | BWt | 44 900 | BWa | 70 000 | kg |
| Conversion factor (mg->kg) | CF | 1 00E-06 | CF | 1 00E-06 | CF | 1 00E-06 | kg/mg |
| Conversion factor (m2->cm2) | CFs | 1 00E+04 | CFs | 1 00E+04 | CFs | 1 00E+04 | cm2/m2 |
| Number of Days near Celotex | DayCc | 350 | DayCt | 350 | DayCa | 350 | day/yr |
| Exposure Frequency - outdoor | EFOc | 184 | EFOt | 177 | EFOa | 167 | day/yr |
| Exposure Frequency - indoor | EFIc | 186 | EFIt | 173 | EFIa | 183 | day/yr |
| Exposure Duration | EDc | 6 000 | EDt | 11 000 | EDa | 1 000 | yr |
| Days per Year | dpy | 365 | dpy | 365 | dpy | 365 | day/yr |
| Years in Lifetime | Lifetime | 70 | Lifetime | 70 | Lifetime | 70 | yr |
| Averaging time | AT | 25550 | AT | 25550 | AT | 25550 | day |
| Average Daily Dose(lifetime) - oral | OADDIc | 6 01E-05 | OADDIt | 1 13E-05 | OADDIa | 7 09E-06 | mg/(kg*day) |
| Total Average Daily Dose - oral | | | | | OADD | 7 03E-06 | mg/(kg*day) |
| Average Daily Dose(lifetime) - dermal | DADDIc | 3 01E-05 | DADDIt | 2 41E-05 | DADDIa | 1 96E-05 | mg/(kg*day) |
| Total Average Daily Dose - dermal | | | | | DADD | 6 65E-06 | mg/(kg*day) |
| Cancer Slope Factor - ingestion | CSF | 7.30 | CSF | 7 30 | CSF | 7 30 | (kg*day)/mg |
| Incremental Lifetime Cancer Risk - oral | | | | | ILCRo | 5 13E-05 | |
| Incremental Lifetime Cancer Risk - dermal | | | | | ILCRd | 4 85E-05 | |
| Incremental Lifetime Cancer Risk - Total | | | | | ILCR | 9 98E-05 | |

Table 11
Estimated Deterministic RBCG
*RME (HEE) - Celotex + AlliedSignal*

| Oral Ingestion | | Toggle | CT Toggle 0 ~~Median~~ | HEE Toggle 1 ~~95th percentile~~ | Point Estimate | Units |
|---|---|---|---|---|---|---|
| BaPeq - outdoor soil | SoilBaPeq | | | | | mg/kg |
| BaPeq - indoor dust | DustBaPeq | | | | | mg/kg |
| Transfer Coefficient | TC | | **0.42** | **0.42** | **0.42** | - |
| Absorption Adjustment Factor - oral | AAFo | 1 | **0.27** | **0.27** | | - |
| Soil Ingestion Rate - child | SIRc | 1 | 62 | 200 | | mg/d |
| Soil Ingestion Rate - teen | SIRt | 1 | 31 | 100 | | mg/d |
| Soil Ingestion Rate - adult | SIRa | 1 | 31 | 100 | | mg/d |
| Conversion factor (mg->kg) | CF | | | | 1 00E-06 | kg/mg |
| Conversion factor (m2->cm2) | CFs | | | | 1 00E+04 | cm2/m2 |
| Body Weight - child | BWc | 1 | 14 7 | 16 4 | | kg |
| Body Weight - teen | BWt | 1 | 42 5 | 44 9 | | kg |
| Body Weight - adult | BWa | 1 | 71 | 70 | | kg |
| Exposure Frequency - outdoor - child | EFOc | 1 | 164 | 164 | | d |
| Exposure Frequency - indoor - child | EFIc | 1 | 186 | 186 | | d |
| Exposure Frequency - outdoor - teen | EFOc | 1 | 177 | 177 | | d |
| Exposure Frequency - indoor - teen | EFIc | 1 | 173 | 173 | | d |
| Exposure Frequency - outdoor - adult | EFOc | 1 | 167 | 167 | | d |
| Exposure Frequency - indoor - adult | EFIc | 1 | 183 | 183 | | d |
| Days per Year | dpy | | | | 350 | d/y |
| Exposure Duration - child | EDc | 1 | 3 2 | 6 | | y |
| Exposure Duration - teen | EDt | 1 | 0 2 | 11 | | y |
| Exposure Duration - adult | EDa | 1 | 0 1 | 1 | | y |
| Years in Lifetime | Lifetime | | | | 70 | y |

| Dermal Contact | | | ~~Median~~ | ~~95th percentile~~ | Point Estimate | Units |
|---|---|---|---|---|---|---|
| BaPeq - outdoor soil | SoilBaPeq | | | | | mg/kg |
| BaPeq - indoor dust | DustBaPeq | | | | | mg/kg |
| Transfer Coefficient | TC | | **0.42** | **0.42** | **0.42** | - |
| Absorption Adjustment Factor - dermal | AAFd | 1 | **0.02** | **0.02** | | - |
| Soil Adherence Rate | SAR | 1 | 0 181 | 1 | | mg/(cm2•d) |
| Dust Adherence Rate | DAR | 1 | 0 045 | 0 2 | | mg/(cm2•d) |
| Skin Surface Area - child | SAc | 1 | 0 63 | 0 73 | | m2 |
| Skin Surface Area - teen | SAt | 1 | 1 34 | 1 5 | | m2 |
| Skin Surface Area - adult | SAa | 1 | 1 8 | 2 | | m2 |
| Fraction of Skin Area Exposed | Frc | | | | 0 25 | - |
| Conversion factor (mg->kg) | CF | | | | 1 00E-06 | kg/mg |
| Conversion factor (m2->cm2) | CFs | | | | 1 00E+04 | cm2/m2 |
| Body Weight - child | BWc | 1 | 14 7 | 16 4 | | kg |
| Body Weight - teen | BWt | 1 | 42 50 | 44 9 | | kg |
| Body Weight - adult | BWa | 1 | 71 | 70 | | kg |
| Exposure Frequency - outdoor - child | EFOc | 1 | 164 | 164 | | d |
| Exposure Frequency - indoor - child | EFIc | 1 | 186 | 186 | | d |
| Exposure Frequency - outdoor - teen | EFOt | 1 | 177 | 177 | | d |
| Exposure Frequency - indoor - teen | EFIt | 1 | 173 | 173 | | d |
| Exposure Frequency - outdoor - adult | EFOa | 1 | 167 | 167 | | d |
| Exposure Frequency - indoor - adult | EFIa | 1 | 183 | 183 | | d |
| Days per Year | dpy | | | | 350 | d/y |
| Exposure Duration - child | EDc | 1 | 3 2 | 6 | | y |
| Exposure Duration - teen | EDt | 1 | 0 2 | 11 | | y |
| Exposure Duration - adult | EDa | 1 | 0 1 | 1 | | y |
| Years in Lifetime | Lifetime | | | | 70 | y |

Table 12
Estimated Deterministic RBCG
Default RME - US EPA Region V

| Variable | Child age: 1 through 6 yr | | Teen age: 7 through 17 yr | | Adult age: 18 through 70 yr | | Units |
|---|---|---|---|---|---|---|---|
| BaPeq - outdoor soil | SoilBaPeq | 1 930 | SoilBaPeq | 1 930 | SoilBaPeq | 1 930 | mg/kg BaPeq |
| Transfer Coefficient | TC | 0 420 | TC | 0 420 | TC | 0 420 | dimensionless |
| BaPeq - indoor dust | DustBaPeq | 0 811 | DustBaPeq | 0 811 | DustBaPeq | 0 811 | |
| Soil Ingestion Rate | SIRc | 200 000 | SIRt | 200 000 | SIRa | 100 000 | mg/day |
| Absorption Adjustment Factor - oral | AAFo | 0 900 | AAFo | 0 900 | AAFo | 0 900 | dimensionless |
| Average Daily Dose (day) - oral -outdoor | OADDOc | 2 32E-05 | OADDOt | 7.72E-06 | OADDOa | 2 48E-06 | mg/(kg•day) |
| Average Daily Dose (day) - oral - indoor | OADDIc | 9 73E-06 | OADDIt | 3 24E-06 | OADDIa | 1 04E-06 | mg/(kg•day) |
| Soil Adherence Rate | SAR | 1 000 | SAR | 1 000 | SAR | 1 000 | mg/(cm2•d) |
| Dust Adherence Rate | DAR | 1.000 | DAR | 1 000 | DAR | 1 000 | mg/(cm2•d) |
| Skin Surface Area | SAc | 0.730 | SAt | 1 500 | SAa | 2.000 | cm2 |
| Fraction of Skin Area Exposed - outdoor | FrcOc | 0 250 | FrcOt | 0 250 | FrcOa | 0 250 | - |
| Fraction of Skin Area Exposed - indoor | FrcIc | 0 250 | FrcIt | 0 250 | FrcIa | 0 250 | - |
| Absorption Adjustment Factor - dermal | AAFd | 0.150 | AAFd | 0 150 | AAFd | 0 150 | dimensionless |
| Average Daily Dose (day) - dermal - outdoor | DADDOc | 3 52E-05 | DADDOt | 2 41E-05 | DADDOa | 2 07E-05 | mg/(kg•day) |
| Average Daily Dose (day) - dermal - indoor | DADDIc | 1 48E-05 | DADDIt | 1 01E-05 | DADDIa | 8 69E-06 | mg/(kg•day) |
| Body Weight | BWc | 15 000 | BWt | 45 000 | BWa | 70 000 | kg |
| Conversion factor (mg->kg) | CF | 1 00E-06 | CF | 1 00E-06 | CF | 1 00E-06 | kg/mg |
| Conversion factor (m2->cm2) | CFs | 1 00E+04 | CFs | 1 00E+04 | CFs | 1 00E+04 | cm2/m2 |
| Number of Days near Celotex | DayCc | 350 | DayCt | 350 | DayCa | 350 | day/yr |
| Exposure Frequency - outdoor | EFOc | 350 | EFOt | 350 | EFOa | 350 | day/yr |
| Exposure Frequency - indoor | EFIc | 0 | EFIt | 0 | EFIa | 0 | day/yr |
| Exposure Duration | EDc | 6 000 | EDt | 11 000 | EDa | 13 000 | yr |
| Days per Year | dpy | 365 | dpy | 365 | dpy | 365 | day/yr |
| Years in Lifetime | Lifetime | 70 | Lifetime | 70 | Lifetime | 70 | yr |
| Averaging time | AT | 25550 | AT | 25550 | AT | 25550 | day |
| Average Daily Dose(lifetime) - oral | OADDIc | 2 22E-05 | OADDIt | 7 40E-06 | OADDIa | 2 38E-06 | mg/(kg•day) |
| Total Average Daily Dose - oral | | | | | OADD | 3 51E-06 | mg/(kg•day) |
| Average Daily Dose(lifetime) - dermal | DADDIc | 3 38E-05 | DADDIt | 2 31E-05 | DADDIa | 1 98E-05 | mg/(kg•day) |
| Total Average Daily Dose - dermal | | | | | DADD | 1 02E-05 | mg/(kg•day) |
| Cancer Slope Factor - ingestion | CSF | 7 30 | CSF | 7 30 | CSF | 7 30 | (kg•day)/mg |
| Incremental Lifetime Cancer Risk - oral | | | | | ILCRo | 2 56E-05 | |
| Incremental Lifetime Cancer Risk - dermal | | | | | ILCRd | 7 46E-05 | |
| Incremental Lifetime Cancer Risk - Total | | | | | ILCR | 1 00E-04 | |

Table 12
Estimated Deterministic RBCG
Default RME - US EPA Region V

| Oral Ingestion | | Toggle | CT Toggle 0 ~~Median~~ | RME Toggle 1 ~~95th percentile~~ | Point Estimate | Units |
|---|---|---|---|---|---|---|
| BaPeq - outdoor soil | SoilBaPeq | | | | | mg/kg |
| BaPeq - indoor dust | DustBaPeq | | | | | mg/kg |
| Transfer Coefficient | TC | | 0 42 | 0 42 | **0.42** | - |
| Absorption Adjustment Factor - oral | AAFo | 1 | 0 9 | 0 9 | | - |
| Soil Ingestion Rate - child | SIRc | 1 | 100 | 200 | | mg/d |
| Soil Ingestion Rate - teen | SIRt | 1 | 100 | 200 | | mg/d |
| Soil Ingestion Rate - adult | SIRa | 1 | 50 | 100 | | mg/d |
| Conversion factor (mg->kg) | CF | | | | 1 00E-06 | kg/mg |
| Conversion factor (m2->cm2) | CFs | | | | 1 00E+04 | cm2/m2 |
| Body Weight - child | BWc | 1 | 15 | 15 | | kg |
| Body Weight - teen | BWt | 1 | 45 | 45 | | kg |
| Body Weight - adult | BWa | 1 | 70 | 70 | | kg |
| Exposure Frequency - outdoor - child | EFOc | 1 | 260 | 350 | | d |
| Exposure Frequency - indoor - child | EFIc | 1 | 90 | 0 | | d |
| Exposure Frequency - outdoor - teen | EFOc | 1 | 260 | 350 | | d |
| Exposure Frequency - indoor - teen | EFIc | 1 | 90 | 0 | | d |
| Exposure Frequency - outdoor - adult | EFOc | 1 | 260 | 350 | | d |
| Exposure Frequency - indoor - adult | EFIc | 1 | 90 | 0 | | d |
| Days per Year | dpy | | 365 | 365 | 350 | d/y |
| Exposure Duration - child | EDc | 1 | 6 | 6 | | y |
| Exposure Duration - teen | EDt | 1 | 2 | 11 | | y |
| Exposure Duration - adult | EDa | 1 | 9 | 13 | | y |
| Years in Lifetime | Lifetime | | 70 | 70 | 70 | y |

| Dermal Contact | | | CT | RME | Point Estimate | Units |
|---|---|---|---|---|---|---|
| BaPeq - outdoor soil | SoilBaPeq | | | | | mg/kg |
| BaPeq - indoor dust | DustBaPeq | | | | | mg/kg |
| Transfer Coefficient | TC | | 0 42 | 0 42 | **0.42** | - |
| Absorption Adjustment Factor - dermal | AAFd | 1 | 0 15 | 0 15 | | - |
| Soil Adherence Rate | SAR | 1 | 0 2 | 1 | | mg/(cm2*d) |
| Dust Adherence Rate | DAR | 1 | 0 2 | 1 | | mg/(cm2*d) |
| Skin Surface Area - child | SAc | 1 | 0 73 | 0 73 | | m2 |
| Skin Surface Area - teen | SAt | 1 | 1 5 | 1 5 | | m2 |
| Skin Surface Area - adult | SAa | 1 | 2 | 2 | | m2 |
| Fraction of Skin Area Exposed | Frc | | | | 0 25 | - |
| Conversion factor (mg->kg) | CF | | | | 1 00E-06 | kg/mg |
| Conversion factor (m2->cm2) | CFs | | | | 1 00E+04 | cm2/m2 |
| Body Weight - child | BWc | 1 | 15 | 15 | | kg |
| Body Weight - teen | BWt | 1 | 45 | 45 | | kg |
| Body Weight - adult | BWa | 1 | 70 | 70 | | kg |
| Exposure Frequency - outdoor - child | EFOc | 1 | 260 | 350 | | d |
| Exposure Frequency - indoor - child | EFIc | 1 | 90 | 0 | | d |
| Exposure Frequency - outdoor - teen | EFOt | 1 | 260 | 350 | | d |
| Exposure Frequency - indoor - teen | EFIt | 1 | 90 | 0 | | d |
| Exposure Frequency - outdoor - adult | EFOa | 1 | 260 | 350 | | d |
| Exposure Frequency - indoor - adult | EFIa | 1 | 90 | 0 | | d |
| Days per Year | dpy | | 365 | 365 | 350 | d/y |
| Exposure Duration - child | EDc | 1 | 6 | 6 | | y |
| Exposure Duration - teen | EDt | 1 | 2 | 11 | | y |
| Exposure Duration - adult | EDa | 1 | 9 | 13 | | y |
| Years in Lifetime | Lifetime | | 70 | 70 | 70 | y |

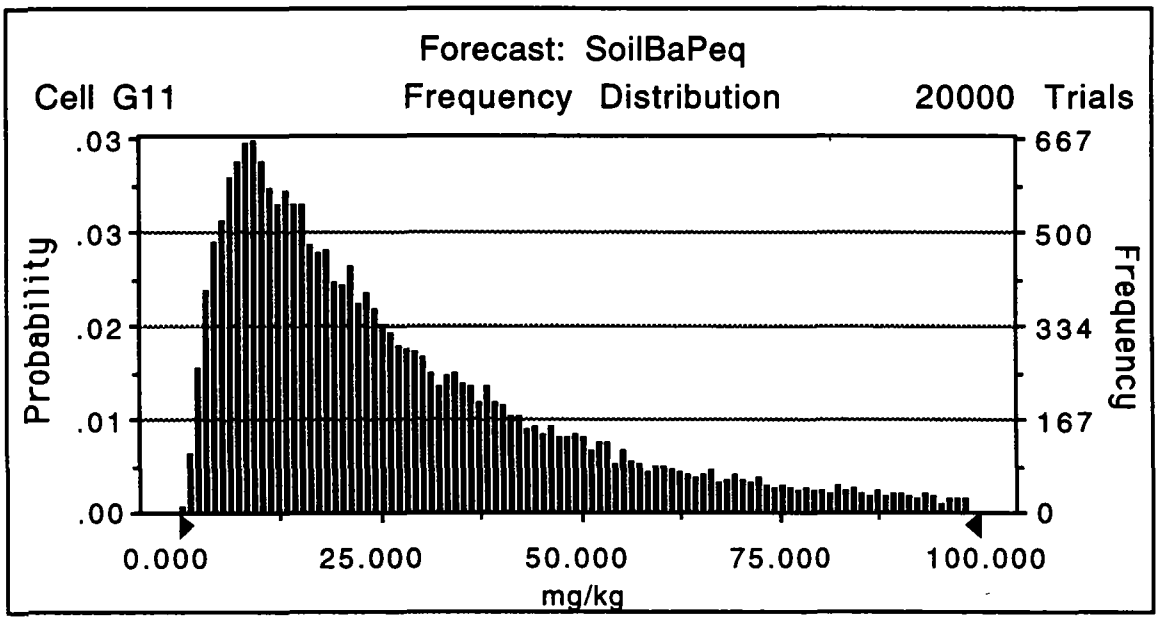
# Tx.Prob.6.SYLK

Crystal Ball® Simulation
Started on Thu, May 30, 1996 at 19:24:07
Stopped on Thu, May 30, 1996 at 20:41:50

**Forecast:** SoilBaPeq **Cell:** G11

Summary: Certainty Level is 100.00% based on Entire Range
Certainty Range is from -∞ to ∞ mg/kg
Display Range is from 0.000 to 100.000 mg/kg
Entire Range is from 0.473 to 99.474 mg/kg
After 20,000 Trials, the Std. Error of the Mean is 0.15

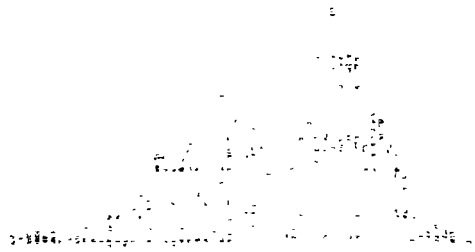
| Statistics: | Display Range | Entire Range |
|---|---|---|
| Trials | 20,000 | 20,000 |
| Percent of Other | 100.00 | 100.00 |
| Mean | 27.118 | 27.118 |
| Median | 20.843 | 20.843 |
| Mode | 7.500 | 7.500 |
| Standard Deviation | 21.120 | 21.120 |
| Variance | 446.063 | 446.063 |
| Skewness | 1.23 | 1.23 |
| Kurtosis | 4.02 | 4.02 |
| Coeff. of Variability | 77.88 | 77.88 |
| Range Width | 100.000 | 99.001 |
| Range Minimum | 0.000 | 0.473 |
| Range Maximum | 100.000 | 99.474 |
| Mean Std. Error | 0.15 | 0.15 |



Percentiles for Entire Range (mg/kg):

| Percentile | SoilBaPeq |
|---|---|
| 0% | 0.473 |
| 10% | 6.355 |
| 20% | 9.470 |
| 30% | 12.781 |
| 40% | 16.409 |
| 50% | 20.844 |
| 60% | 26.058 |
| 70% | 33.256 |
| 80% | 42.569 |
| 90% | 58.164 |
| 100% | 99.474 |

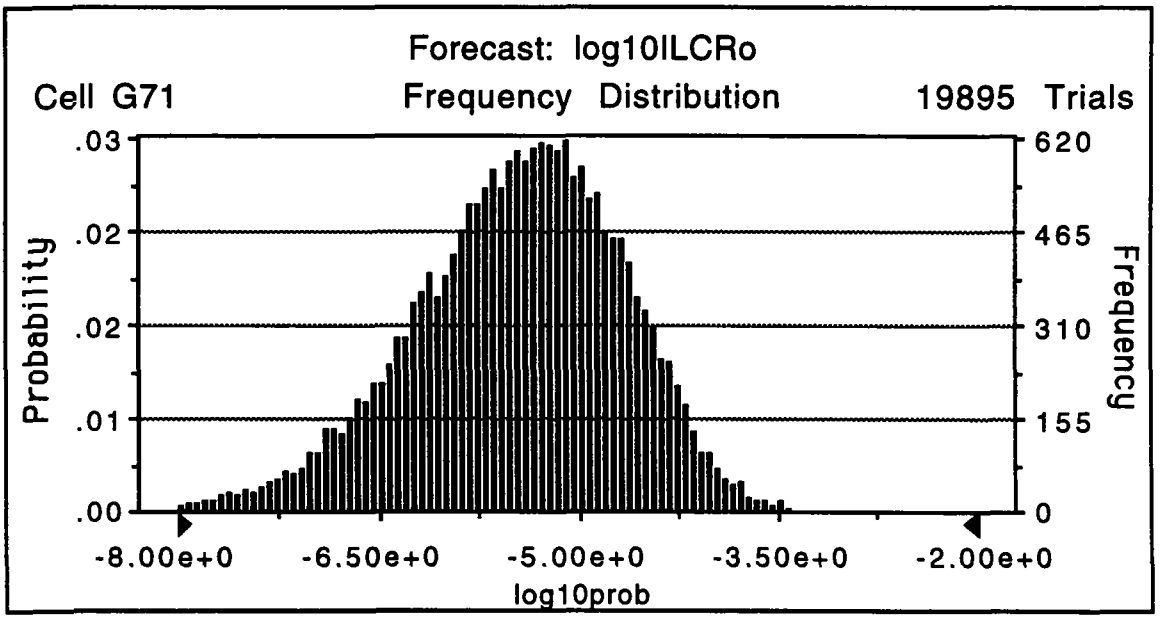
End of Forecast



**Forecast:** log10ILCRo **Cell:** G71

Summary: Certainty Level is 100.00% based on Entire Range
Certainty Range is from -∞ to ∞ log10prob
Display Range is from -8.00E+0 to -2.00E+0 log10prob
Entire Range is from -9.87E+0 to -2.95E+0 log10prob
After 20,000 Trials, the Std. Error of the Mean is 0.01

| Statistics: | Display Range | Entire Range |
|---|---|---|
| Trials | 19,895 | 20,000 |
| Percent of Other | 99.47 | 100.53 |
| Mean | -5.45e+0 | -5.46e+0 |
| Median | -5.39e+0 | (unavailable) |
| Mode | -5.09e+0 | (unavailable) |
| Standard Deviation | 7.84e-1 | 8.11e-1 |
| Variance | 6.15e-1 | 6.58e-1 |
| Skewness | -0.34 | (unavailable) |
| Kurtosis | 2.96 | (unavailable) |
| Coeff. of Variability | -14.40 | -14.85 |
| Range Width | 6.00e+0 | 6.92e+0 |
| Range Minimum | -8.00e+0 | -9.87e+0 |
| Range Maximum | -2.00e+0 | -2.95e+0 |
| Mean Std. Error | 0.01 | 0.01 |




Percentiles for Entire Range (log10prob):

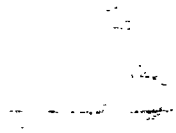
| Percentile | log10ILCRo |
|---|---|
| 0% | -9.87E+0 |
| 10% | -6.52E+0 |
| 20% | -6.12E+0 |
| 30% | -5.83E+0 |
| 40% | -5.60E+0 |
| 50% | -5.40E+0 |

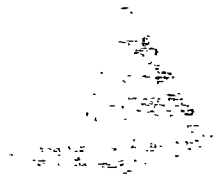
**Forecast:** log10ILCRo (Cont'd)

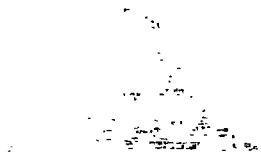
**Cell:** G71

| Percentile | log10ILCRo |
|---|---|
| 60% | -5.20E+0 |
| 70% | -5.00E+0 |
| 80% | -4.77E+0 |
| 90% | -4.48E+0 |
| 100% | -2.95E+0 |

End of Forecast



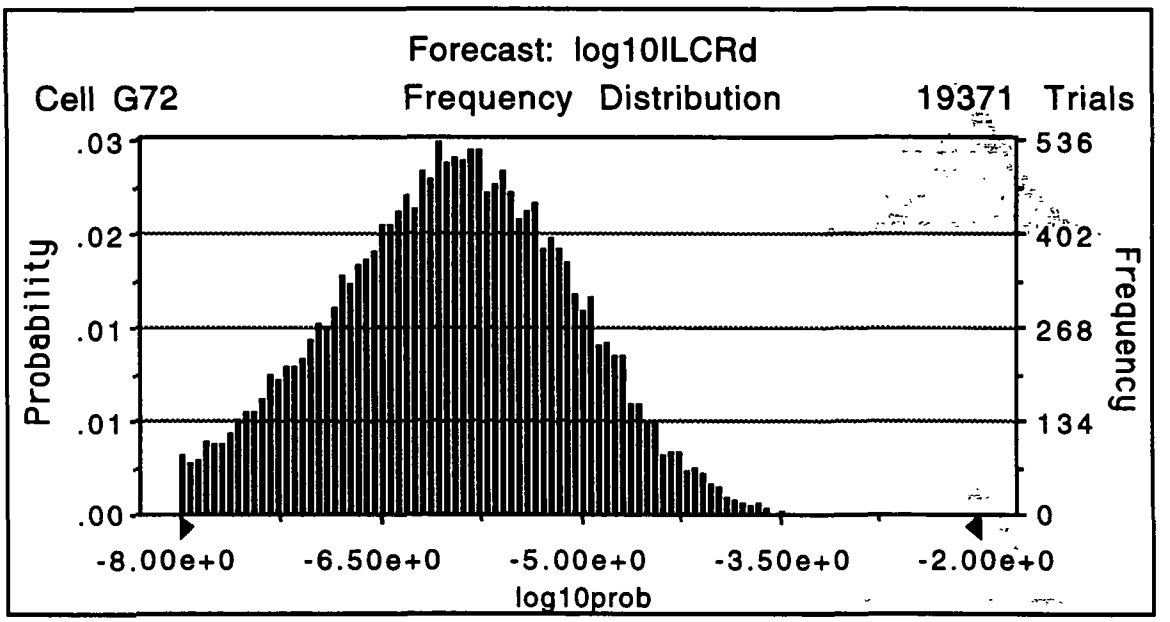




**Forecast:** log10ILCRd **Cell:** G72

Summary: Certainty Level is 100.00% based on Entire Range
Certainty Range is from -∞ to ∞ log10prob
Display Range is from -8.00E+0 to -2.00E+0 log10prob
Entire Range is from -1.13E+1 to -2.75E+0 log10prob
After 20,000 Trials, the Std. Error of the Mean is 0.01

| Statistics: | Display Range | Entire Range |
|---|---|---|
| Trials | 19,371 | 20,000 |
| Percent of Other | 96.86 | 103.25 |
| Mean | -5.95e+0 | -6.03e+0 |
| Median | -5.93e+0 | (unavailable) |
| Mode | -5.83e+0 | (unavailable) |
| Standard Deviation | 8.76e-1 | 9.80e-1 |
| Variance | 7.68e-1 | 9.60e-1 |
| Skewness | -0.04 | (unavailable) |
| Kurtosis | 2.57 | (unavailable) |
| Coeff. of Variability | -14.73 | -16.25 |
| Range Width | 6.00e+0 | 8.55e+0 |
| Range Minimum | -8.00e+0 | -1.13e+1 |
| Range Maximum | -2.00e+0 | -2.75e+0 |
| Mean Std. Error | 0.01 | 0.01 |




Percentiles for Entire Range (log10prob):

| Percentile | log10ILCRd |
|---|---|
| 0% | -1.13E+1 |
| 10% | -7.31E+0 |
| 20% | -6.81E+0 |
| 30% | -6.48E+0 |
| 40% | -6.20E+0 |
| 50% | -5.96E+0 |

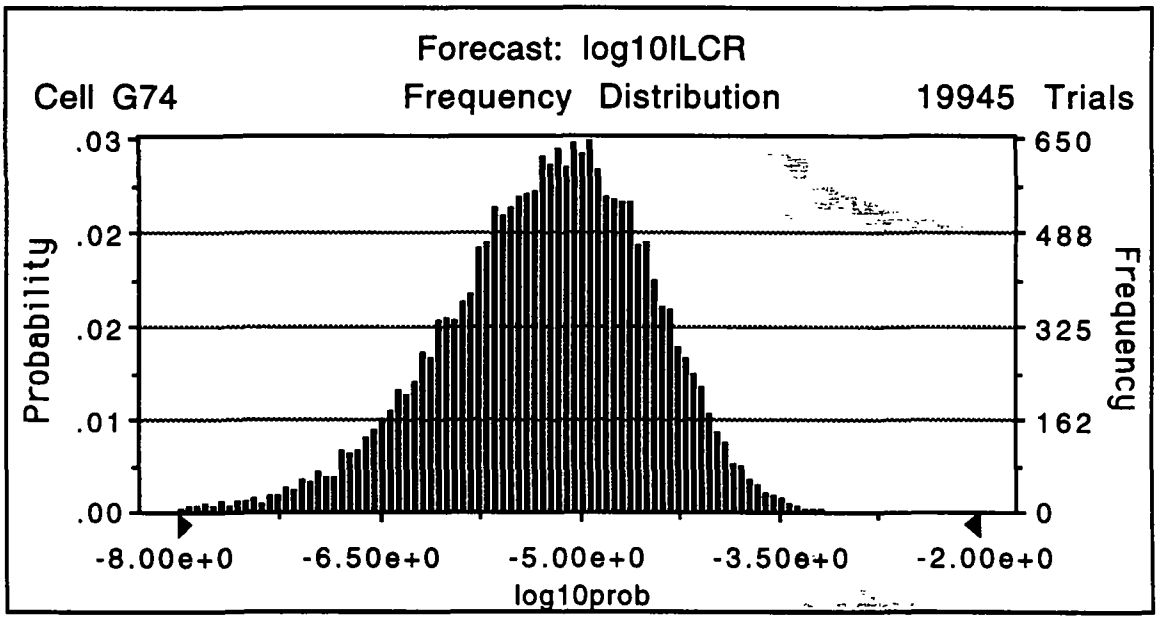
| Percentile | log10ILCRd |
|---|---|
| 60% | -5.73E+0 |
| 70% | -5.48E+0 |
| 80% | -5.20E+0 |
| 90% | -4.83E+0 |
| 100% | -2.75E+0 |

End of Forecast

**Forecast:** log10ILCR **Cell:** G74

Summary: Certainty Level is 100.00% based on Entire Range
Certainty Range is from -∞ to ∞ log10prob
Display Range is from -8.00E+0 to -2.00E+0 log10prob
Entire Range is from -9.64E+0 to -2.75E+0 log10prob
After 20,000 Trials, the Std. Error of the Mean is 0.01

| Statistics: | Display Range | Entire Range |
|---|---|---|
| Trials | 19,945 | 20,000 |
| Percent of Other | 99.72 | 100.28 |
| Mean | -5.24e+0 | -5.25e+0 |
| Median | -5.18e+0 | (unavailable) |
| Mode | -5.05e+0 | (unavailable) |
| Standard Deviation | 7.72e-1 | 7.88e-1 |
| Variance | 5.95e-1 | 6.22e-1 |
| Skewness | -0.41 | (unavailable) |
| Kurtosis | 3.14 | (unavailable) |
| Coeff. of Variability | -14.72 | -15.01 |
| Range Width | 6.00e+0 | 6.89e+0 |
| Range Minimum | -8.00e+0 | -9.64e+0 |
| Range Maximum | -2.00e+0 | -2.75e+0 |
| Mean Std. Error | 0.01 | 0.01 |

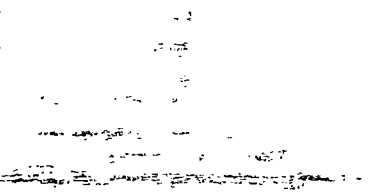



Percentiles for Entire Range (log10prob):

| Percentile | log10ILCR |
|---|---|
| 0% | -9.64E+0 |
| 10% | -6.28E+0 |
| 20% | -5.88E+0 |
| 30% | -5.61E+0 |
| 40% | -5.39E+0 |
| 50% | -5.19E+0 |

| Percentile | log10ILCR |
|---|---|
| 60% | -5.00E+0 |
| 70% | -4.80E+0 |
| 80% | -4.58E+0 |
| 90% | -4.30E+0 |
| 100% | -2.75E+0 |

End of Forecast

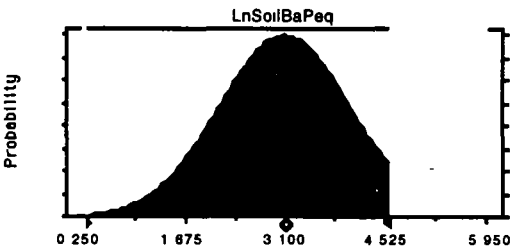


## Assumptions

**Assumption:** LnSoilBaPeq **Cell:** G10

Normal distribution with parameters:
Mean 3.100
Standard Dev. 0.950

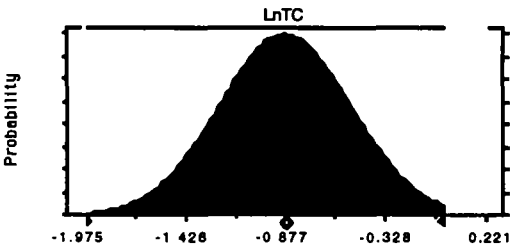
Selected range is from -∞ to 4.600
Mean value in simulation was 2.981




**Assumption:** LnTC **Cell:** G12

Normal distribution with parameters:
Mean -0.877
Standard Dev. 0.366

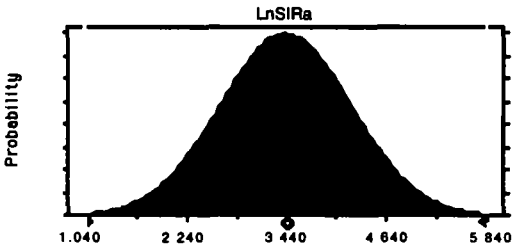
Selected range is from -∞ to 0.000
Mean value in simulation was -0.886




**Assumption:** LnSIRa **Cell:** G17

Normal distribution with parameters:
Mean 3.440
Standard Dev. 0.800

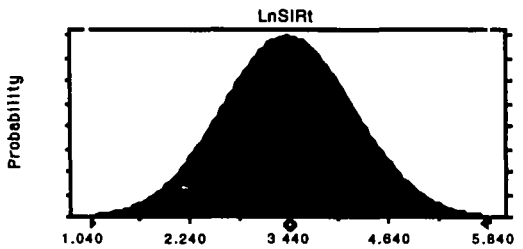
Selected range is from -∞ to ∞
Mean value in simulation was 3.445




**Assumption:** LnSIRt **Cell:** E17

Normal distribution with parameters:
Mean 3.440
Standard Dev. 0.800

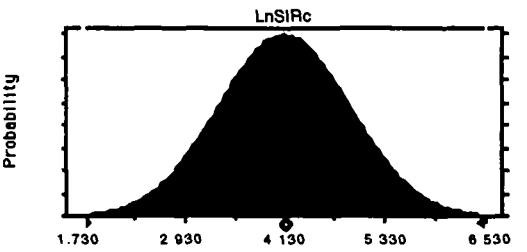
Selected range is from -∞ to ∞
Mean value in simulation was 3.446



**Assumption:** LnSIRc **Cell:** C17

Normal distribution with parameters:
Mean 4.130
Standard Dev. 0.800

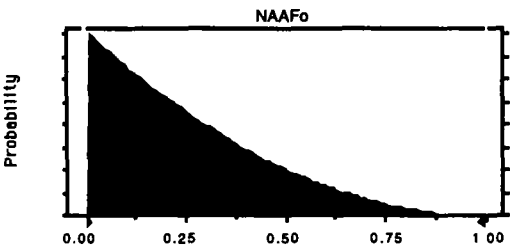
Selected range is from -∞ to ∞
Mean value in simulation was 4.129




**Assumption:** NAAFo **Cell:** G19

Beta distribution with parameters:
Alpha 1.00
Beta 3.00

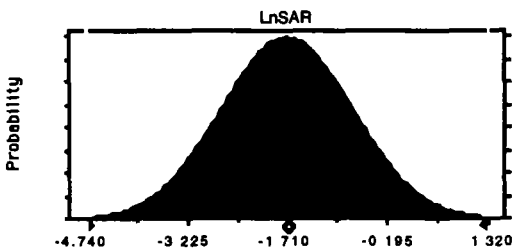
Selected range is from 0.00 to 1.00
Mean value in simulation was 0.25




**Assumption:** LnSAR **Cell:** G25

Normal distribution with parameters:
Mean -1.710
Standard Dev. 1.010

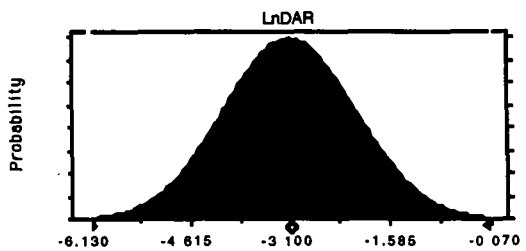
Selected range is from -∞ to ∞
Mean value in simulation was -1.718




**Assumption:** LnDAR **Cell:** G27

Normal distribution with parameters:
Mean -3.100
Standard Dev. 1.010

Selected range is from -∞ to ∞
Mean value in simulation was -3.092



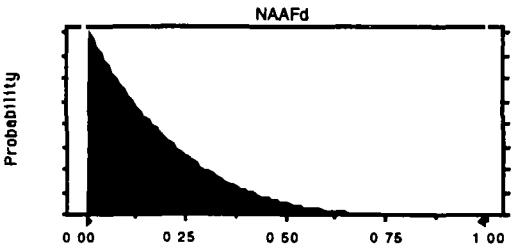

**Assumption:** NAAFd **Cell:** G31

Beta distribution with parameters:
Alpha 1.00
Beta 5.00

Selected range is from 0.00 to 1.00

**Assumption:** NAAFd (Cont'd) **Cell:** G31

Mean value in simulation was 0.17

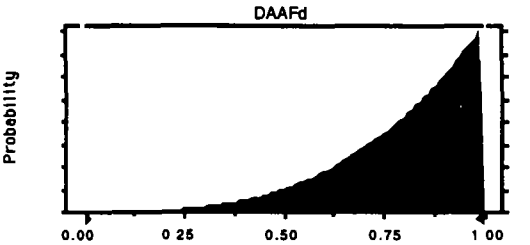



**Assumption:** DAAFd **Cell:** G32

Beta distribution with parameters:

| | |
|---|---|
| Alpha | 4.00 |
| Beta | 1.00 |

Selected range is from 0.00 to 1.00
Mean value in simulation was 0.80

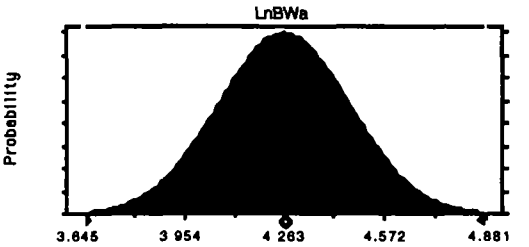



**Assumption:** LnBWa **Cell:** G38

Normal distribution with parameters:

| | |
|---|---|
| Mean | 4.263 |
| Standard Dev. | 0.206 |

Selected range is from -∞ to ∞
Mean value in simulation was 4.264

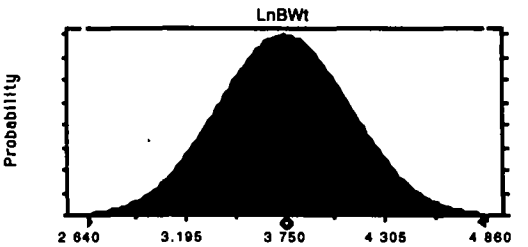



**Assumption:** LnBWt **Cell:** E38

Normal distribution with parameters:

| | |
|---|---|
| Mean | 3.750 |
| Standard Dev. | 0.370 |

Selected range is from -∞ to ∞
Mean value in simulation was 3.755




**Assumption:** LnBWc **Cell:** C38

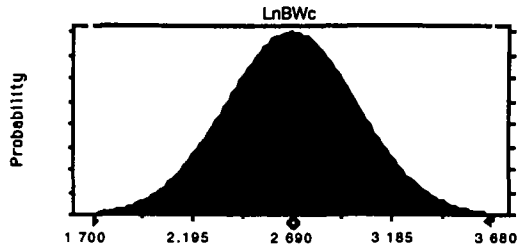
Normal distribution with parameters:

| | |
|---|---|
| Mean | 2.690 |
| Standard Dev. | 0.330 |

Selected range is from -∞ to ∞

**Assumption:** LnBWc (Cont'd) **Cell:** C38

Mean value in simulation was 2.690

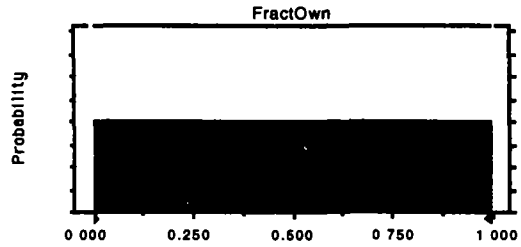



**Assumption:** FractOwn **Cell:** G49

Uniform distribution with parameters:

| | |
|---|---|
| Minimum | 0.000 |
| Maximum | 1.000 |

Selected range is from 0.000 to 1.000
Mean value in simulation was 0.503

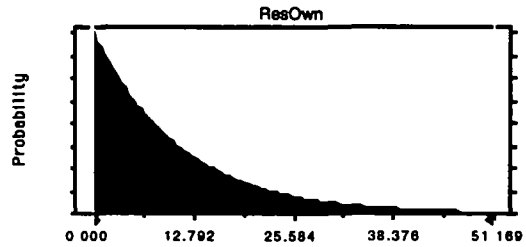



**Assumption:** ResOwn **Cell:** G50

Exponential distribution with parameters:

| | |
|---|---|
| Rate | 0.090 |

Selected range is from 0.000 to ∞
Mean value in simulation was 11.062

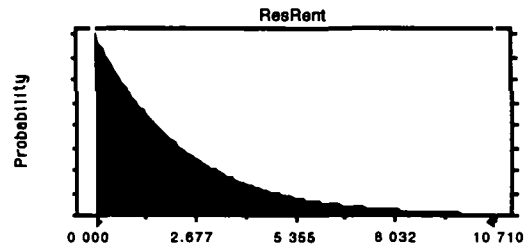



**Assumption:** ResRent **Cell:** G51

Exponential distribution with parameters:

| | |
|---|---|
| Rate | 0.430 |

Selected range is from 0.000 to ∞
Mean value in simulation was 2.346




**Assumption:** EFOc **Cell:** C45

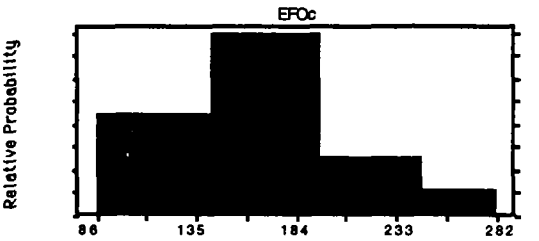
Custom distribution with parameters:

| | | | Relative Probability |
|---|---|---|---|
| Continuous range | 86 to | 143 | 0.30 |
| Continuous range | 143 to | 196 | 0.50 |
| Continuous range | 196 to | 246 | 0.15 |
| Continuous range | 246 to | 282 | 0.05 |

**Assumption:** EFOc (Cont'd) **Cell:** C45

Total Relative Probability 1.00

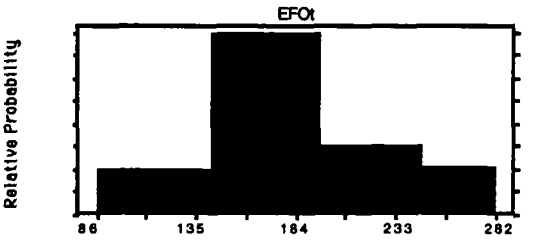
Mean value in simulation was 165




**Assumption:** EFOt **Cell:** E45

| Custom distribution with parameters: | | | Relative Probability |
|---|---|---|---|
| Continuous range | 86 to | 143 | 0.15 |
| Continuous range | 143 to | 196 | 0.55 |
| Continuous range | 196 to | 246 | 0.20 |
| Continuous range | 246 to | 282 | 0.10 |
| Total Relative Probability | | | 1.00 |

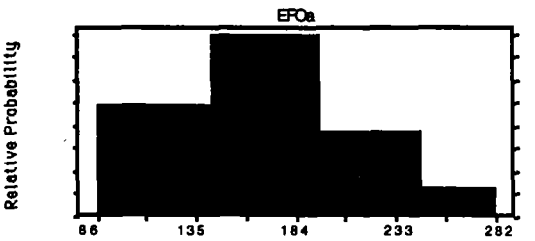
Mean value in simulation was 181




**Assumption:** EFOa **Cell:** G45

| Custom distribution with parameters: | | | Relative Probability |
|---|---|---|---|
| Continuous range | 86 to | 143 | 0.30 |
| Continuous range | 143 to | 196 | 0.45 |
| Continuous range | 196 to | 246 | 0.20 |
| Continuous range | 246 to | 282 | 0.05 |
| Total Relative Probability | | | 1.00 |

Mean value in simulation was 168




End of Assumptions

# Appendix A

# Background Concentrations of Benzo(a)pyrene in Soil Samples Near the Industrial Property

25 October 1996

Alceon Corporation
PO Box 382669
Harvard Square Station
Cambridge, MA 02238-2669

617-864-4300

# Appendix A

## Background Concentrations of Benzo(a)pyrene in Surface Soil Samples Near the Industrial Property

In this Appendix, we present the measurements of the concentrations of carcinogenic polycyclic aromatic hydrocarbons (cPAHs) -- as expressed in terms of benzo(a)pyrene equivalents (mg/kg BaPeq, equivalent to ppm) -- for surface soil samples representing "urban background concentrations" for the site in Chicago, IL.

As shown in Table A-1, we have organized the surface soil samples into three groups.

- Group A - the 7 surface soil samples taken by Illinois Environmental Protection Agency (IEPA) during the early 1990s from residential properties located to the west of Kedzie Blvd. The US Environmental Protection Agency has nominated these samples as representative of urban background conditions, although no formal sampling plan nor QA/QC program for these samples has been provided.

- Group B - the 2 surface soil samples taken by IEPA during 1991 and 1992 from Douglas Park, more than 2,500 ft north of the Celotex property. The US Environmental Protection Agency has proposed these samples as representative of urban background conditions, although no formal sampling plan nor QA/QC program for these samples has been provided.

- Group C - the 40 surface soil samples taken by ERM - North Central (ERM) during 1995 from residential properties located in a band between the radii of 1,500 ft and 2,500 ft in Sectors 1 and 8 to the north of the Celotex property. Using powerful statistical methods, Dr. Louis Anthony Cox, Jr. of Cox Associates has demonstrated that the spatial patterns of these concentrations are unrelated -- with 95 percent confidence -- to any airborne cPAHs that may have emanated at any time from the industrial property (See Appendix B.) Thus, these 40 samples, collected under a stringent sampling plan with strong QA/QC provisions, now provide a strong statistical population for samples representing urban background concentrations.

Taken together, these 49 samples provide a statistical population of measurements against which other populations of measurements may be compared using nonparametric tests such as the Wilcoxon Rank Sum test or the Kolmogorov-Smirnov test.

# LogNormal Probability Plot for 49 Samples

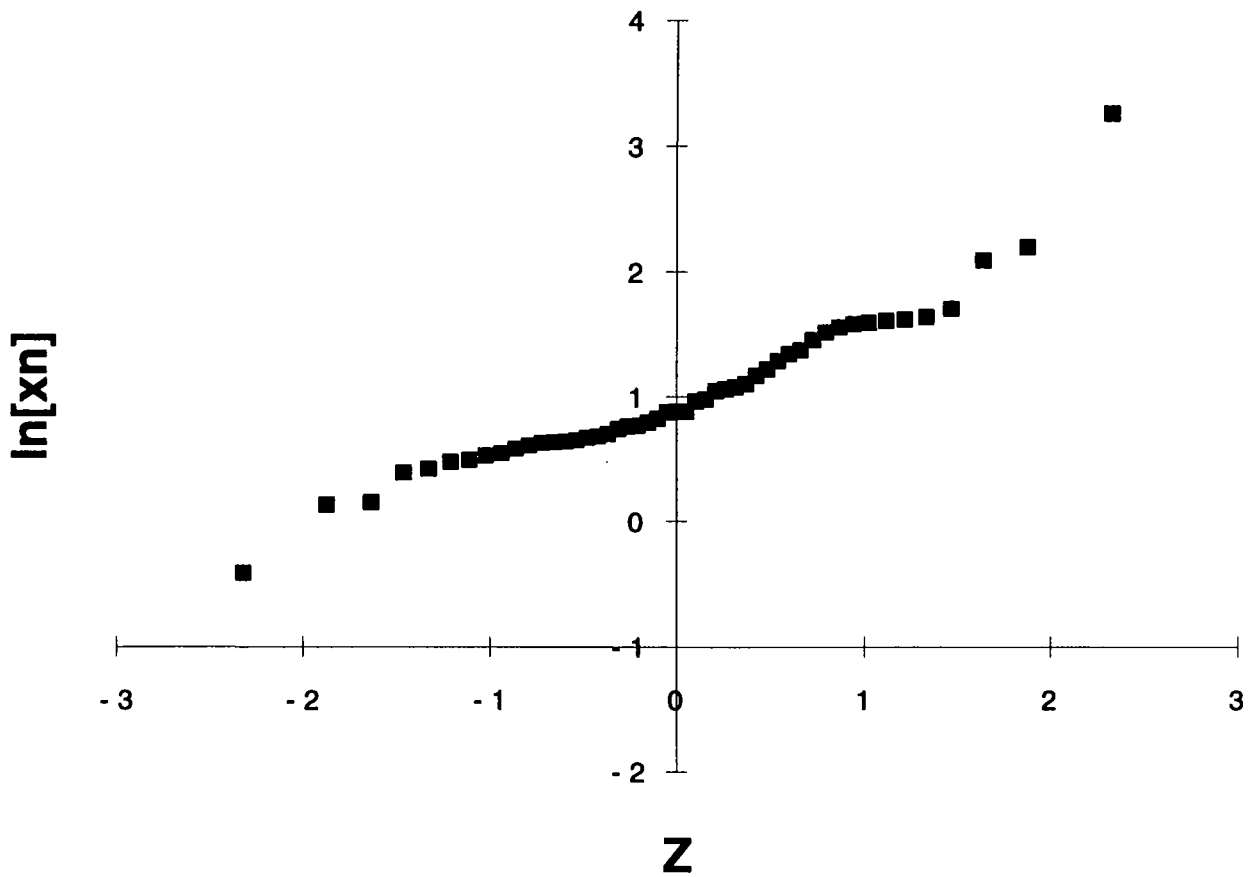


Table A-1
Soil Samples Representing "Urban Background Concentrations"

| Group A | | Group B | | Group C | |
|---|---|---|---|---|---|
| **Map ID** | **BaPeq mg/kg** | **Map ID** | **BaPeq mg/kg** | **Map ID** | **BaPeq mg/kg** |
| ID236 | 1.2 | ID200 | 0.7 | ID25 | 4.5 |
| ID237 | 1.8 | ID229 | 1.9 | ID26 | 4.9 |
| ID238 | 9.0 | | | ID32 | 4.0 |
| ID239 | 1.9 | | | ID14 | 2.9 |
| ID240 | 1.7 | | | ID15 | 2.9 |
| ID243 | 1.8 | | | ID16 | 4.7 |
| ID244 | 1.9 | | | ID17 | 3.2 |
| | | | | ID18 | 5.0 |
| | | | | ID19 | 4.3 |
| | | | | ID20 | 2.4 |
| | | | | ID21 | 2.8 |
| | | | | ID22 | 3.6 |
| | | | | ID23 | 2.7 |
| | | | | ID24 | 2.2 |
| | | | | ID71 | 2.3 |
| | | | | ID72 | 1.7 |
| | | | | ID73 | 3.4 |
| | | | | ID74 | 1.6 |
| | | | | ID75 | 3.8 |
| | | | | ID76 | 1.5 |
| | | | | ID77 | 2.0 |
| | | | | ID78 | 1.5 |
| | | | | ID79 | 2.4 |
| | | | | ID80 | 2.1 |
| | | | | ID81 | 2.4 |
| | | | | ID82 | 2.6 |
| | | | | ID83 | 1.6 |
| | | | | ID84 | 2.2 |
| | | | | ID85 | 1.1 |
| | | | | ID86 | 2.0 |
| | | | | ID87 | 2.0 |
| | | | | ID88 | 1.9 |
| | | | | ID27 | 2.2 |
| | | | | ID28 | 5.5 |
| | | | | ID29 | 4.9 |
| | | | | ID30 | 26.0 |
| | | | | ID31 | 8.1 |
| | | | | ID89 | 3.0 |
| | | | | ID90 | 5.1 |
| | | | | ID91 | 5.0 |

Notes:

A detection limit of 0.5 ppm is assumed when no detection limit is reported.

For Background data, a detection limit of 0.5 ppm is assumed when no value whatsoever is reported.

# Appendix B

# Estimating the Spatial Extent of Site-Related Contamination

25 October 1996

Cox Associates
503 Franklin Street
Denver, CO 80218

303-541-6043

Appendix B

# ESTIMATING THE SPATIAL EXTENT OF SITE-RELATED CONTAMINATION

## CONCEPTS FOR SPATIAL DATA ANALYSIS

To assess the spatial extent of the human health risk that might potentially be associated with the Celotex property, it is useful determine whether there is any distance from the fenced Celotex property beyond which there is no association between soil concentrations of carcinogenic PAHs (measured as B(a)P equivalents) and distance from the property. If soil concentrations are statistically independent of distance from the Celotex property at all locations more than a certain distance, d, from the Celotex property, then these locations may be considered to be "background" locations for the purposes of quantifying the effects of contamination due to the Celotex property. (The distance d may be different in different directions, due to asymmetries in the wind rose or transport mechanisms) In other words, **background locations** may be defined as locations where there is no evidence of any contamination from the Celotex property. A first task is to identify whether there are such background locations.

## METHODS FOR SPATIAL DATA ANALYSIS

Even small levels of contamination from the Celotex property can prevent a location from being classified as background. Therefore, it is desirable to use techniques that are not sensitive to the absolute magnitudes of soil contamination. To detect possible small but consistent additions to contamination above background levels, it is useful to apply statistical techniques that examine the spatial pattern of concentrations and that seek to identify where (if anywhere) they stop being related to distance from the Celotex property. Such statistical methods are ordinal: they test whether closer proximity to the property is associated with higher soil concentrations, without regard for the absolute magnitude of the concentrations.

The following statistical methodology was used to identify distances beyond which there is no significant association between soil concentrations and distance from the Celotex property.

*Step 1. Choose a coordinate system and compute distances from the center of the Celotex property to each location.* The original geographic coordinates ("Easting" and "Northing") were translated to assign the coordinates (0, 0) to the ERM map origin. (This origin is close to Map ID #209, near the center of the fenced location.) Next, the distance from the origin to each other location was computed. (Of course, these distances do not depend on the translation of axes.) These distances provide the key information for testing whether distance from the Celotex property (defined as distance from the origin) is associated with concentration.

*Note:* The analyses were repeated using locations 202, 209, 210, 211, 212, 219, and 220 as the origin. None of the conclusions changes based on which exact location within the fenced area is taken as the origin.

*Step 2: Hypothesis testing.* The formal statistical description of this step is as follows. For different distances, d, around the Celotex property, test the following null hypothesis:

H0: "For locations more than d feet from the origin, there is no association between distance from the Celotex property and soil concentration."

against the alternative hypothesis

H1: "For locations more than d feet from the origin, there is an association between distance from the Celotex property and soil concentration."

Find the smallest distance, d* (if there is one) such that H1 is rejected in favor of H0 (at a p = 5% significance level) for all distances greater than d*. This test was performed using the Spearman rank correlation coefficient (Siegel, 1956). It was checked by repeating the analysis using Kendall's Tau (Siegel, 1956) rather than Spearman's rank correlation coefficient to quantify the strength of association between distance and concentration.

Informally, the logic of this step is as follows. For any given radius, d, all locations more than d feet from the Celotex property are sorted in increasing order of their distances from the property (i.e., from the origin). Next, it is checked whether the locations with the highest concentrations tend to occur disproportionately often toward the top of the sorted list (i.e., whether higher concentrations tend to be observed more often than would be expected based on chance alone among locations closer to the origin.) This calculation is made using the statistical theory of pairwise ordinal associations. The relevant statistical theory and computational techniques (e.g., Spearman's rank correlation coefficient and Kendall's Tau) have been widely accepted and applied as part of mainstream nonparametric statistical theory for more than forty years. They provide quantitative measures of the ordinal association between concentrations and distances from the origin. More importantly for the purposes of this analysis, they allow quantitative calculation of the probability of observing by chance alone (i.e., in the absence of any true association) an association between high concentrations and low distances at least as strong as the association actually observed in the sample data. If this probability (the "p-value", or significance level, of the test) is small enough, then the hypothesis of no association is rejected in favor of the hypothesis that there is an association.

The results from this step can be summarized by a table giving the p-value of the observed association between distance and concentration among locations at least d feet from the origin (and outside the fenced Celotex property), for different values of d.

*Technical note:* The null hypotheses for increasing values of d actually form a nested family of hypotheses, tested using smaller and smaller subsets of the same data (namely, locations at increasing distances from the origin). In principle, this could raise complications due to the problem known as "multiple comparisons" (arising from testing multiple hypotheses based on the same data; see e.g., Bechoffer, 1995). In practice, the number of data points available for hypothesis testing is relatively large (it is 39 for the largest radius considered, 1500 feet), so that the hypothesis tests have adequate power to

detect even relatively weak associations and the problem of multiple comparisons does not threaten the validity of the conclusions.

*Step 3: Refinements and validation of the hypothesis tests.* The hypothesis-testing procedure just described was refined by applying it to different angular sectors (e.g., to one quadrant at a time). In addition, whenever it was concluded that locations more than d feet from the Celotex property in a certain direction (angular sector) were not affected by the Celotex property, then this conclusion was cross-checked by computing the average gradient (the direction of steepest increase in concentrations; see Benveniete et al., 1995) from the sampled data. If the conclusion is correct, the gradient directions computed from subsets of locations more than d from the origin should show no tendency to point toward the origin more than they point away from it. These refinements and validation tests generally confirmed the conclusions based on the Spearman's rank correlation analysis, so they are not reported on further here.

*Step 4: Spatial data analysis based on concentric rings.* The preceding steps were discussed with the EPA (especially, Dr. Arthur Lubin), who contributed substantial advice and suggested additional tests and improvements. Dr. Lubin pointed out that to the extent that the spatial data tend to be distributed at approximately the same distance from the origin (so that radial variation toward or away from the origin is small compared to transverse variation), the Spearman's rank correlation test and similar methods may be powerless to detect an association between distance from the origin and concentration, even if such an association exists (or would exist and be revealed if there were adequate radial variation in the sampling plan). To overcome this difficulty, we reanalyzed the data using an entirely different statistical logic. First, the data were subdivided into 10 concentric rings at increasing distances around the origin. Next, the frequency distributions of concentrations among locations within different rings were compared. These comparisons were carried out using the Kolmogorov-Smirnov (K-S) nonparametric test, (DeGroot, 1975), based on a recommendation from Dr. Lubin. This test has the advantage that it avoids any need to make potentially controversial assumptions about the specific parametric forms of the concentration distributions. The use of concentric rings allows the K-S test to examine absolute differences among distributions, instead of relying on ordinal tests such as the Spearman's rank

correlation test. Thus, it provides an alternative statistical logic for investigating the same issues originally explored using the K-S test. As it turned out, the concentric ring analysis using the K-S test confirmed the results of the Spearman's analysis, so the details of the analysis will not be repeated here. However, we note that the robustness of the conclusions was double-checked by repeating the concentric ring analysis using 12 rings instead of 10, making a different assignment of locations to rings, and shifting the origin slightly. In addition, the concentric ring analysis was applied to individual quadrants. None of these variations changed the conclusions, suggesting that the main results of the spatial data analysis are not sensitive to the exact details of the analytic procedure (i.e., they are "robust" to reasonable changes in the statistical data analysis techniques applied).

*Step 5: Validation of conclusions using linear regression and nonparametric ("loess") nonlinear regression and classification tree analyses.* To further validate our conclusions, we repeated the statistical analyses using ordinary linear regression of concentration (and also log of concentration) against distance (in different quadrants and in the whole sample) instead of the Spearman's rank correlation approach. We also used two more sophisticated techniques from modern computational statistics: classification tree analysis (now included in S-PLUS and other advanced statistical computing and artificial intelligence packages) and nonlinear, nonparametric smoothing ("loess" regression), also included in S-PLUS. These more advanced methods confirmed the main results from the simpler analyses, as reported next.

## RESULTS OF SPATIAL DATA ANALYSES

### Results of Exploratory Analysis Using Spearman's Rank Correlation Test

Table 1 shows the main results of the analysis examining the association between distance and "concentration" (measured in B(a)P equivalents) among locations at various distances outside the fence. When locations 1,000 feet or less from the Celotex property are included, the association between distance and concentration is highly significant (p values less than 0.01).

When only locations 1,200 feet or more from the origin are considered, however, the association becomes much less strong -- less than the 5% significance level often used as a default level for rejecting the hypothesis of association. Thus, Table 1 suggests that there is a break between 1,100 and 1,200 feet from the Celotex property, with a significant relation between concentration and distance for locations inside the 1,100-foot radius, but not for neighborhoods outside the 1,100-foot radius.

**TABLE 1:** DISTANCE AND SOIL CONCENTRATION ARE NOT SIGNIFICANTLY ASSOCIATED BEYOND 1100 FEET FROM THE CELOTEX PROPERTY

| Distance in feet (from ID #209) | p-value of association |
|---|---|
| 800 feet | < 0.000001 |
| 900 | 0.000026 |
| 950 | 0.00025 |
| 1000 | 0.00077 |
| 1100 | 0.00377 |
| 1200 | 0.076 |
| 1300 | 0.19 |
| 1400 | 0.08 |
| 1500 | 0.19 |

The above table offers some possible evidence of an association at distances beyond 1,100 feet, although it is not significant at the $p = 0.05$ level. To obtain a more thorough understanding of the data, it is useful to examine the concentration-distance relation in different directions around the origin. The results are shown in Table 2.

TABLE 2: THE ASSOCIATION BETWEEN CONCENTRATION AND DISTANCE APPEARS TO EXTEND FURTHEST IN THE NORTHEAST QUADRANT

| | NW Quadrant | | SW Quadrant | | NE Quadrant | |
|---|---|---|---|---|---|---|
| distance in feet | p-value | N | p | N | p | N |
| 800 | 0.0072 | 48 | 0.021 | 20 | 0.000002 | 29 |
| 900 | 0.035 | 46 | " | " | 0.0000045 | 26 |
| 1000 | 0.14 | 44 | 0.047 | 15 | " | " |
| 1100 | " | " | 0.28 | 13 | " | " |
| 1200 | 0.15 | 39 | 0.41 | 10 | 0.0058 | 22 |
| 1500 | 0.51 | 25 | 0.22 | 5 | 0.016 | 21 |

In this table, each quadrant with data (excluding the southeast quadrant, for which there were no soil samples) has two columns of numbers: (i) The p-values associated with the Spearman's rank correlation (between concentration and distance from the origin) for locations outside the Celotex property fence line and more (less???) than the specified distance from the origin; and (ii) The number of locations falling in each of these concentric subsets. Ditto marks indicate distance ranges in which there are no sample data points, so that increasing the distance does not change the results.

These data suggest that in the northwest and southwest quadrants, the association between concentration and distance may become insignificant between 900 and 1100 feet. In the northeast quadrant, the association is significant at distances out to 1500 feet and beyond. (Only two samples were taken between 1500 feet and 1700 feet from the origin, and neither of them was taken from the northeast quadrant, so the exact pattern of concentration vs. distance between 1500 and 1700 feet cannot be determined.) Note that this hypothesis-testing procedure does not provide an exact boundary between significant and non-significant associations, since there is some noise in the p-values based on the sample data, and since data are scarce in the distance range from 1500 to 1700 feet. Also, the fact that the data sets considered are concentric introduces a multiple hypothesis testing problem. However, it is clear from these data that the northeast quadrant deserves additional analysis

and may be a greater source of potential concern than the other two quadrants, based on the sample data.

These analyses, which we regard as exploratory but useful, show that to the west of the Celotex property, *there is no evidence of a significant association between concentration and distance for locations more than 1500 feet from the Celotex property.* [There is about a 46% probability that the observed association among locations beyond 1500 feet, or a stronger one, would occur by chance. This conclusion was double-checked using Kendall's Tau (Siegel, 1956), which applies a different statistical logic based on the similarity of rankings of locations by distance from the origin and by concentration. The resulting p-level was 0.46, indicating that the observed degree of association between concentration and distance for locations beyond 1500 feet is no stronger than would be expected to occur by chance alone in the absence of any true association.] In contrast, *there is strong evidence of a negative association between distance and concentration (i.e., higher concentrations occur at smaller distances from the Celotex property) among locations less than 900 feet from the Celotex property.* Between about 1000 and 1400 feet, the evidence is ambiguous and conclusions are uncertain due to sampling variability ("noise") in the sample data. To the northeast, there is evidence of a significant association at distances out to 1500 feet and beyond. This association is worth additional examination.

Thus, if the goal is to identify locations for which one can be confident that there is significant contamination associated with (distance from) the Celotex property, then one might choose locations inside the 1000-foot radius (or further out in the northeast). If the goal is to identify locations for which one can be quite confident that there is no significant association with distance from the Celotex property, then one could choose neighborhoods outside the 1500 foot radius (and further out in the northeast quadrant).

Results of Other Tests: Ring Analysis, Linear and Nonlinear Regression, Tree Analysis

More detailed analyses using the K-S test to compare the concentration distributions in "rings" at different distances from the origin established the following key results:

1. *Distance is not significantly associated with concentration among all locations more than 1,200 feet from the origin.*

2. *In the northeast quadrant specifically, distance is not significantly associated with concentration among all locations more than 1,500 feet from the origin.* (There is only one data point between 1200 and 1500 feet from the origin in the northeast quadrant, so that possible association between distance and concentration over this interval cannot be determined from the available data.)

These findings are further supported by the "loess" nonparametric nonlinear regression, which suggests that the negative association between distance from the origin and concentration disappears between 1200 and 1500 feet from the origin in the northeast quadrant and in the whole data set. Simple linear regression of concentration or log-concentration against distance leads to similar conclusions. For example, the log-concentration regression coefficient for DISTANCE when all data points in the northeast quadrant (out to a mile) and more than 1100 feet from the origin are considered is highly significant ($p = 0.00015$). At distances of 1200 feet and more, the regression coefficient for distance becomes insignificant ( $p \geq 0.2$). This is consistent with the hypothesis that there is no significant association between concentration and distance beyond 1,200 feet from the origin, even in the northeast quadrant.

Finally, the classification tree analysis, which uses a very different approach (minimizing classification entropy) from any of the other methods considered, reaches very compatible conclusions. When asked to discover rules for predicting concentration in B(a)P equivalents for locations outside the Celotex fence, based on the available sample data, this method automatically

determines that locations more than 1193 feet from the origin tend to have significantly lower concentrations than locations closer to the origin. It automatically clusters the locations into the following four rings:

- Less than 470 feet from the origin (mean concentration = 30.5).
- 471 - 763 feet from the origin (mean concentration = 27)
- 764 - 1193 feet from the origin (mean concentration = 10.4)
- More than 1193 feet from the origin (mean concentration = 3.4).

Thus, to a close approximation, this method also leads to the conclusion that locations more than 1,200 feet from the origin (in any direction) may be pooled together and treated as "background" locations for purposes of statistical analysis.

## DISCUSSION OF SPATIAL DATA ANALYSIS RESULTS

The analysis reported here suggests that 1,200 feet may be used as a statistically supported boundary between "background" locations (not affected by the Celotex property in any statistically detectable or observable way) and locations that might plausibly have been affected. Given the scarcity of data points between 1,200 feet and 1,500 feet from the origin, and the desire to be health-protective in the absence of relevant information, it appears reasonable to treat 1,500 feet as a useful outer bound on the distance at which contamination from the Celotex property affects soil sample concentrations. At the other extreme, it is almost certain that locations less than about 900 to 1000 feet from the Celotex property have concentrations that tend to increase as one moves toward the property. A reasonable compromise between these extremes might be at a radius of about 1,100 feet for locations west of the Celotex property (i.e., in the northwest or southwest quadrants) and at a radius of about 1300 feet to its northeast, as distances at which the hypothesis of contamination associated with the Celotex property cannot be either proved or disproved based on the available data.

The frequency distribution of concentrations among "background" locations (defined conservatively as locations more than 1500 feet from the origin) is well-approximated (according to a K-S test) by a log-normal

distribution with mean 1.0 and variance 0.4. The range of variation in a sample of size 40 spans more than an order of magnitude, from 0.65 to over 26. This has important implications for risk management. Any remediation planning effort that seeks to make the concentrations at locations near the Celotex property indistinguishable from (or at least as clean as) the concentrations at background locations must recognize this variability. For example, the goal of cleanup activities might be to reduce the distribution of concentrations among locations potentially affected by the Celotex property so that it is indistinguishable from the empirical distribution of concentrations at background locations. In this case, the decision of which locations to address first and when to stop should reflect the inherent variability in the concentration distribution among background locations.

The concentration distribution among locations less than 1,000 feet from the Celotex property but outside the fence is well approximated (again according to a K-S test) by a log-normal distribution with mean 2.8 and variance 0.6. The range of values is from 2.1 to 61, almost a 30-fold range. This suggests that there may be substantial gains to health protection to be achieved by focusing on locations in the upper tail of this distribution. Moreover, there is substantial overlap between the concentration distributions of "foreground" locations (e.g., those less than 1,000 feet from the origin but outside the fence) and "background" locations (e.g., those more than 1,500 feet from the origin). Focusing on locations in the upper tail of the foreground locations can help to minimize the overlap between concentrations at properties that are addressed and concentrations at background locations that are not addressed.

REFERENCES

Bechhofer, R.E., T.J. Santner, and D.M. Goldsman, 1995. *Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons.* Wiley, New York.

Benveniste, A., M. Metevier, and P. Priouret, 1990. *Adaptive Algorithms and Stochastic Approximations.* Springer-Verlag, New York.

DeGroot, M.H., 1975. *Introduction to Probability and Statistics.* Addison-Wesley, Reading, MA.

Gilbert, R.O., 1987. *Statistical Methods for Environmental Pollution Monitoring.* Von Nostrand Reinhold, New York.

Siegel, S., 1956. *Nonparametric Statistics,* especially Chapter 9. McGraw-Hill, New York.

# Appendix C

# Toxicity Profile for Benzo(a)pyrene

# from US EPA's IRIS Database

# Downloaded on 28 November 1995

25 October 1996

Alceon Corporation
PO Box 382669
Harvard Square Station
Cambridge, MA 02238-2669

617-864-4300

# Benzo(a)pyrene

1 - IRIS
NAME - Benzo[a]pyrene (BaP)
RN - 50-32-8
IRSN - 133
DATE - 941102
UPDT - 11/02/94, 1 field
STAT - Oral RfD Assessment (RDO) no data
STAT - Inhalation RfC Assessment (RDI) no data
STAT - Carcinogenicity Assessment (CAR) on-line 11/01/94
STAT - Drinking Water Health Advisories (DWHA) no data
STAT - U.S. EPA Regulatory Actions (EXSR) on-line 01/01/92
IRH - 08/01/89 REFS Bibliography on-line
IRH - 01/01/92 CAR Carcinogen assessment noted as pending change
IRH - 01/01/92 EXSR Regulatory actions updated
IRH - 04/01/92 CAR Summary revised; oral quantitative section added
IRH - 04/01/92 CREF Carcinogen assessment references revised
IRH - 05/01/92 CARDR Work group review and verification date corrected
IRH - 07/01/92 CAR Text revised in NOTE
IRH - 07/01/92 CARO Range of slope factors corrected
IRH - 07/01/92 CARO Slope factor and risks corrected
IRH - 07/01/92 CARO Data table heading corrected
IRH - 07/01/92 CARO Slope factor corrected; last paragraph
IRH - 07/01/92 CARDR Secondary contact changed
IRH - 09/01/93 CAR Carcinogenicity assessment noted as pending change
IRH - 09/01/93 CARDR Work group review date added
IRH - 12/01/93 CREF Reference revised - U.S. EPA, 1991b
IRH - 02/01/94 CARDR Primary contact's phone number changed
IRH - 03/01/94 CAR Pending change note removed; no change
IRH - 03/01/94 CARDR Work group review date added
IRH - 07/01/94 CARDR Work group review date added
IRH - 11/01/94 CARO Slope factor clarified; changed O to "0"
RLEN - 25760
SY - BaP
SY - Benzo[a]pyrene
SY - BENZO(d,e,f)CHRYSENE
SY - 3,4-BENZOPIRENE
SY - 3,4-BENZOPYRENE
SY - 6,7-BENZOPYRENE
SY - BENZO(a)PYRENE
SY - 3,4-BENZPYREN
SY - 3,4-BENZPYRENE
SY - 3,4-BENZ(a)PYRENE
SY - BENZ(a)PYRENE
SY - 3,4-BENZYPYRENE
SY - BP
SY - 3,4-BP
SY - B(a)P
SY - RCRA WASTE NUMBER U022

---

RDO - NO DATA

---

RDI - NO DATA
CAREV-
o CLASSIFICATION : B2; probable human carcinogen
o BASIS FOR CLASSIFICATION : Human data specifically linking benzo[a]pyrene (BAP) to a carcinogenic effect are lacking. There are, however, multiple animal studies in many species demonstrating BAP to be carcinogenic following administration by numerous routes. BAP has produced positive results in numerous genotoxicity assays. NOTE: At the June 1992 CRAVE Work Group meeting, a revised risk estimate for benzo[a]pyrene was verified (see Additional Comments for Oral Exposure). This

section provides information on three aspects of the carcinogenic risk assessment for the agent in question; the U.S. EPA classification, and quantitative estimates of risk from oral exposure and from inhalation exposure. The classification reflects a weight-of-evidence judgment of the likelihood that the agent is a human carcinogen. The quantitative risk estimates are presented in three ways. The slope factor is the result of application of a low-dose extrapolation procedure and is presented as the risk per (mg/kg)/day. The unit risk is the quantitative estimate in terms of either risk per ug/L drinking water or risk per ug/cu.m air breathed. The third form in which risk is presented is a drinking water or air concentration providing cancer risks of 1 in 10,000 or 1 in 1,000,000. The Carcinogenicity Background Document provides details on the rationale and methods used to derive the carcinogenicity values found in IRIS. Users are referred to the Oral Reference Dose (RfD) and Reference Concentration (RfC) sections for information on long-term toxic effects other than carcinogenicity.

o HUMAN CARCINOGENICITY DATA :

Inadequate. Lung cancer has been shown to be induced in humans by various mixtures of polycyclic aromatic hydrocarbons known to contain BAP including cigarette smoke, roofing tar and coke oven emissions. It is not possible, however, to conclude from this information that BAP is the responsible agent.

---

o ANIMAL CARCINOGENICITY DATA :

Sufficient. The animal data consist of dietary, gavage, inhalation, intratracheal instillation, dermal and subcutaneous studies in numerous strains of at least four species of rodents and several primates. Repeated BAP administration has been associated with increased incidences of total tumors and of tumors at the site of exposure. Distant site tumors have also been observed after BAP administration by various routes. BAP is frequently used as a positive control in carcinogenicity bioassays.

BAP administered in the diet or by gavage to mice, rats and hamsters has produced increased incidences of stomach tumors. Neal and Rigdon (1967) fed BAP (purity not reported) at concentrations of 0, 1, 10, 20, 30, 40, 45, 50, 100 and 250 ppm in the diets of male and female CFW-Swiss mice. The age of the mice ranged from 17-180 days old and the treatment time from 1-197 days; the size of the treated groups ranged from 9 to 73. There were 289 mice (number of mice/sex not stated) in the control group. No forestomach tumors were reported in the 0-, 1- and 10-ppm dose groups. The incidence of forestomach tumors in the 20-, 30-, 40-, 45-, 50-, 100- and 250-ppm dose groups were 1/23, 0/37, 1/40, 4/40, 23/34, 19/23 and 66/73, respectively. The authors felt that the increasing tumor incidences were related to both the concentration and the number of doses administered. Historical control forestomach tumor data are not available for CFW-Swiss strain mice. In historical control data from a related mouse strain, SWR/J Swill, the forestomach tumor incidence rate was 2/268 and 1/402 for males and females, respectively (Rabstein et al., 1973).

Brune et al., (1981) fed 0.15 mg/kg BAP (reported to be "highly pure") in the diet of 32 Sprague-Dawley rats/sex/group either every 9th day or 5 times/week. These treatments resulted in annual average doses of 6 or 39 mg/kg, respectively. An untreated group of 32 rats/sex served as the control. Rats were treated until moribund or dead; survival was similar in all groups. Histologic examinations were performed on each rat. The combined incidence of tumors of the forestomach, esophagus and larynx was 3/64, 3/64 and 10/64 in the control group, the group fed BAP every 9th day and the group fed BAP 5

times/week, respectively. A trend analysis showed a statistically significant tendancy for the proportion of animals with tumors of the forestomach, esophagus or larynx to increase steadily with dose (Knauf and Rice, 1992).

As part of the same study, Brune et al. (1981) administered BAP ("highly pure") orally to Sprague-Dawley rats by caffeine gavage. The rats were treated until moribund or dead; all rats were subjected to terminal histopathologic examination. Gavaged rats were divided into 3 dose groups of 32 rats/sex/group; the groups received 0.15 mg/kg per gavage either every 9th day (Group A), every 3rd day (Group B) or 5 times per week (Group C); these treatments resulted in annual average doses of 6, 18 or 39 mg/kg, respectively. Untreated and gavage (5 times/week) controls (32 rats/sex/group) were included. The median survival times for the untreated control group; the gavage control group; and groups A, B and C were 129, 102, 112, 113 and 87 weeks, respectively. The survival time of Group C was short compared with controls and may have precluded tumor formation (Knauf and Rice, 1992). The combined tumor incidence in the forestomach, esophagus and larynx was 3/64, 6/64, 13/64, 26/64 and 14/64 for the untreated control group, gavage control group, group A, group B and group C, respectively. There was a statistically significant association between the dose and the proportions of rats with tumors of the forestomach, esophagus or larynx. This association is not characterized by a linear trend. The linearity was affected by the apparently reduced tumor incidence that is seen in the high-dose group (Knauf and Rice, 1992).

Intratracheal instillation and inhalation studies in guinea pigs, hamsters and rats have resulted in elevated incidences of respiratory tract and upper digestive tract tumors (U.S. EPA, 1991a). Male Syrian golden hamsters (24/group) were exposed by inhalation to 0, 2.2, 9.5 or 46.5 mg BAP/cu.m in a sodium chloride aerosol (Thyssen et al., 1981). (Greater than 99% of the particles had diameters between 0.2 and 0.5 um.) For the first 10 weeks of the study, the hamsters were exposed to BAP daily for 4.5 hours/day; thereafter, daily for 3 hours/day. Animals dying within the first year of the study were replaced; the effective number of hamsters in the control, low-, mid- and high-dose groups was 27, 27, 26 and 25, respectively. (The total time of treatment, although over 60 weeks, was not stated.) During the first 10 weeks, animals in the 3 dose groups reportedly lost weight. After week 10, however, the body weights in all groups were similar until week 60 when the body weights of hamsters in the high-dose group decreased and the mortality increased significantly. The incidence of respiratory tract tumors (including tumors of the nasal cavity, larynx and trachea) in the control, low-, mid- and high-dose groups was 0/27, 0/27, 9/26 and 13/25, respectively; the incidences of upper digestive tract tumors (including tumors of the pharynx, esophagus and forestomach) were 0/27, 0/27, 7/26 and 14/25, respectively. Trend analysis for incidences of both respiratory tract tumors and upper gastrointestinal tract tumors showed a statistically significant tendancy for the proportion of animals with either tumor type to increase steadily with increased dose (Knauf and Rice, 1992).

Intraperitoneal BAP injections have caused increases in the number of injection site tumors in mice and rats (reviewed in U.S. EPA, 1991a). Subcutaneous BAP injections have caused increases in the number of injection site tumors in mice, rats, guinea pigs, hamsters and some primates (IARC, 1983; U.S. EPA, 1991a). BAP is commonly used as a positive control in many dermal application bioassays and has been shown to cause skin tumors in mice, rats, rabbits and guinea pigs. BAP is both an initiator and a complete carcinogen in mouse skin (IARC, 1983). Increased incidences of distant site tumors have also been reported in animals as a consequence of dermal BAP exposure (reviewed in U.S. EPA, 1991a).

BAP has also been reported to be carcinogenic in animals when administered by the following routes: i.v.; transplacentally; implantation in the stomach wall, lung, renal parenchyma and brain; injection into the renal pelvis; and vaginal painting (U.S. EPA, 1991a).

---

o SUPPORTING DATA :

Benzo[a]pyrene has been shown to cause genotoxic effects in a broad range

of prokaryotic and mammalian cell assay systems (U.S. EPA, 1991a). In prokaryotes, BAP tested positive in DNA damage assays and in both reverse and forward mutation assays. In mammalian cell culture assays, BAP tested positive in DNA damage assays, forward mutation assays, chromosomal effects assays and cell transformation assays.

---

---

CARO -

o CLASSIFICATION : B2; probable human carcinogen

o BASIS FOR CLASSIFICATION : Human data specifically linking benzo[a]pyrene (BAP) to a carcinogenic effect are lacking. There are, however, multiple animal studies in many species demonstrating BAP to be carcinogenic following administration by numerous routes. BAP has produced positive results in numerous genotoxicity assays. NOTE: At the June 1992 CRAVE Work Group meeting, a revised risk estimate for benzo[a]pyrene was verified (see Additional Comments for Oral Exposure). This section provides information on three aspects of the carcinogenic risk assessment for the agent in question; the U.S. EPA classification, and quantitative estimates of risk from oral exposure and from inhalation exposure. The classification reflects a weight-of-evidence judgment of the likelihood that the agent is a human carcinogen. The quantitative risk estimates are presented in three ways. The slope factor is the result of application of a low-dose extrapolation procedure and is presented as the risk per (mg/kg)/day. The unit risk is the quantitative estimate in terms of either risk per ug/L drinking water or risk per ug/cu.m air breathed. The third form in which risk is presented is a drinking water or air concentration providing cancer risks of 1 in 10,000 or 1 in 1,000,000. The Carcinogenicity Background Document provides details on the rationale and methods used to derive the carcinogenicity values found in IRIS. Users are referred to the Oral Reference Dose (RfD) and Reference Concentration (RfC) sections for information on long-term toxic effects other than carcinogenicity.

o ORAL SLOPE FACTOR : 7.3E+0 per (mg/kg)/day

o DRINKING WATER UNIT RISK : 2.1E-4 per (ug/L)

o DOSE EXTRAPOLATION METHOD : Risk estimate based on a geometric mean of four slope

o RISK/WATER CONCENTRATIONS :

Drinking Water Concentrations at Specified Risk Levels:

| Risk Level | Concentration |
|---|---|
| E-4 (1 in 10,000) | 5E-1 ug/L |
| E-5 (1 in 100,000) | 5E-2 ug/L |
| E-6 (1 in 1,000,000) | 5E-3 ug/L |

---

o ORAL DOSE-RESPONSE DATA :

Tumor Type -- forestomach, squamous cell papillomas and carcinomas
Test Animals -- CFW mice, sex unknown
Route -- oral, diet

Reference -- Neal and Rigdon, 1967

a) Conditional upper bound two-stage model with terms for promotion (modification of Moolgavkar-Venson-Knudson, generalized forms of two-stage model)

| Administered Dose (ppm) | Tumor Incidence |
|---|---|
| 0 | 0/289 |
| 1 | 0/25 |
| 10 | 0/24 |
| 20 | 1/23 |
| 30 | 0/37 |
| 40 | 1/40 |
| 45 | 4/40 |
| 50 | 24/34 |
| 100 | 19/23 |
| 250 | 66/73 |

Tumor Type -- squamous cell carcinoma of the forestomach
Test Animals -- SWR/J Swill mice
Route -- oral, diet
Reference -- Rabstein et al., 1973

| Administered Dose (ppm) | Tumor Incidence |
|---|---|
| 0 | 2/268* male |
| 0 | 1/402* female |

*See additional comments concerning the use of control data from other studies that utilized similar mouse strains.

b) Same data as above. Upper bound estimate by extrapolation from 10% response point to background of empirically fitted dose-response curve. (Procedure using two-stage model described in (a)).

c) Same data as above except the additional 2 control groups (Rabstein et al., 1973) were excluded. Generalized Weibull-type dose-response model.

d) Tumor Type -- forestomach, larynx and esophagus, papillomas and carcinomas (combined). Linearized Multistage Model, Extra Risk.

Test Animals -- Sprague-Dawley rats, males and females
Route -- oral, diet
Reference -- Brune et al., 1981

| Dose (mg/kg diet/year) | Tumor Incidence |
|---|---|
| 0 | 3/64 |
| 6 | 3/64 |
| 39 | 10/64 |

---

o ADDITIONAL COMMENTS :

NOTE: The range of oral slope factors calculated was: 4.5E+0 to 11.7E+0 per (mg/kg)/day.

At the June 1992 CRAVE Work Group meeting, it was noted that an error had been made in the 1991 document "Dose-Response Analysis of Ingested Benzo[a]pyrene" which is quoted in the Drinking Water Criteria Document for PAH. In the calculation of the doses in the Brune et al. (1981) study it was erroneously concluded that doses were given in units of mg/year, whereas it was in fact mg/kg/year. When the doses are corrected the slope factor is correctly calculated as 11.7 per (mg/kg)/day, as opposed to 4.7 per (mg/kg)/day as reported in the Drinking Water Criteria Document. The correct

range of slope factors is 4.5 to 11.7 per (mg/kg)/day, with a geometric mean of 7.3 per (mg/kg)/day. A drinking water unit risk based on the revised slope factor is 2.1E-4 per (ug/L). Therefore, these values have been changed on IRIS and an Erratum to the Drinking Water Criteria Document is being prepared.

Risk estimates were calculated from two different studies in two species of outbred rodents (Neal and Rigdon, 1967; Brune et al., 1981). These studies have several commonalities including mode of administration, tumor sites, tumor types and the presumed mechanisms of action. The data sets were not combined prior to modeling (the preferred approach) because they employed significantly dissimilar protocols.

The geometric mean from several slope factors, each considered to be of equal merit, was used to calculate a single unit risk. These four slope factor estimates span less than a factor of three and each is based on an acceptable, but less-than-optimal, data set. Each estimate is based on a low-dose extrapolation procedure which entails the use of multiple assumptions and default procedures.

Clement Associates (1990) fit the Neal and Rigdon (1967) data to a two-stage dose response model. In this model the transition rates and the growth rate of preneoplastic cells were both considered to be exposure-dependent. (The functional form for the dose-dependence of preneoplastic cell growth rate was simple saturation.) A term to permit the modeling of BAP as its own promoter was also included. Historical control stomach tumor data from a related, but not identical, mouse strain, SWR/J Swill (Rabstein et al., 1973) and the CFW Texas colony (Neal and Rigdon, 1967) were used in the modeling. In calculating the lifetime unit risk for humans several standard assumptions were made: mouse food consumption was 13% of its body weight/day; human body weight was assumed to be 70 kg and the assumed body weight of the mouse 0.034 kg. The standard assumption of surface area equivalence between mice and humans was the cube root of 70/0.034. A conditional upper bound estimate was calculated to be 5.9 per (mg/kg)/day (U.S. EPA, 1991a).

A U.S. EPA report (1991b) argued that the upper-bound estimate calculated in Clement Associates (1990) involved the use of unrealistic conditions placed on certain parameters of the equation. Other objections to this slope factor were also raised. The authors of this report used the Neal and Rigdon (1967) data to generate an upper-bound estimate extrapolated linearly from the 10% response point to the background of an empirically fitted dose-response curve (Clement Associates, 1990). Other results, from similar concepts and approaches used for other compounds, suggest that the potency slopes calculated in this manner are comparable to those obtained from a linearized multistage procedure for the majority of the other compounds. The upper bound estimate calculated in U.S. EPA (1991b) is 9.0 per (mg/kg)/day.

The authors of U.S. EPA (1991b) selected a model to reflect the partial lifetime exposure pattern over different parts of the animals' lifetimes. The authors thought that this approach more closely reflected the Neal and Rigdon (1967) regimen. A Weibull-type dose-response model was selected to accommodate the partial lifetime exposure; the upper-bound slope factor calculated from this method was 4.5 per (mg/kg)/day.

Using the dietary portion of the Brune et al. (1981) rat data, a linearized multistage procedure was used to calculate an upper bound slope factor for humans. In the interspecies conversion the assumed human body weight was 70 kg and the rat 0.4 kg. The slope factor calculated by this method was 11.7 per (mg/kg)/day.

---

o DISCUSSION OF CONFIDENCE :

The data are considered to be less than optimal, but acceptable. There are precedents for using multiple data sets from different studies using more than one sex, strain and species; the use of the geometric mean of four slope factors is preferred because it makes use of more of the available data. The use of the geometric means was based on arguments presented in a personal communication (Stiteler, 1991).

---
---
CARI - NO DATA
CARDR-
o CARCINOGENICITY SOURCE :

Source Document -- U.S. EPA, 1991a,b

The 1991 Drinking Water Criteria Document for the polycyclic aromatic hydrocarbons has received agency review.
DOCUMENT
---
o REVIEW DATES : 01/07/87, 12/04/91, 06/03/92, 08/05/93, 02/02/94,
o VERIFICATION DATE : 12/04/91
o EPA CONTACTS :

Robert E. McGaughy / OHEA -- (202)260-5889

Rita Schoeny / OHEA -- (513)569-7544
---
---
HAONE- NO DATA
---
HATEN- NO DATA
---
HALTC- NO DATA
---
HALTA- NO DATA
---
HALIF- NO DATA
---
OLEP - NO DATA
---
ALAB - NO DATA
---
TREAT- NO DATA
---
HADR - NO DATA
---
CAA - NO DATA
---
WQCHU-

Water and Fish Consumption: 2.8E-3 ug/L

Fish Consumption Only: 3.11E-2 ug/L

Considers technological or economic feasibility? -- NO

Discussion -- For the maximum protection from the potential carcinogenic properties of this chemical, the ambient water concentration should be zero. However, zero may not be obtainable at this time, so the recommended criteria represents a E-6 estimated incremental increase of cancer over a lifetime. The values given represent polynuclear aromatic hydrocarbons as a class.

Reference -- 45 FR 79318 (11/28/80)

EPA Contact -- Criteria and Standards Division / OWRS
(202)260-1315 / FTS 260-1315

---
---
WQCAQ-

Freshwater:

Acute LEC -- none
Chronic LEC -- none

Marine:

Acute LEC -- 3.0E+2 ug/L
Chronic LEC -- none

Considers technological or economic feasibility? -- NO

Discussion -- The values that are indicated as "LEC" are not criteria, but are the lowest effect levels found in the literature. LEC's are given when the minimum data required to derive water quality criteria are not available. The values given represent polynuclear aromatic hydrocarbons as a class.

Reference -- 45 FR 79318 (11/28/80)

EPA Contact -- Criteria and Standards Division / OWRS
(202)260-1315 / FTS 260-1315

---
---
---

MCLG -

Value -- 0 mg/L (Proposed,1990)

Considers technological or economic feasibility? -- NO

Discussion -- The proposed MCLG for benzo(a)pyrene is zero based on the evidence of carcinogenic potential (B2).

Reference -- 55 FR 30370 (07/25/90)

EPA Contact -- Health and Ecological Criteria Division / OST /
(202) 260-7571 / FTS 260-7571; or Safe Drinking Water Hotline / (800) 426-4791

---
---

MCL -

Value -- 0.0002 mg/L (Proposed, 1990)

Considers technological or economic feasibility? -- YES

Discussion -- The proposed MCL is equal to the PQL and is associated with a maximum lifetime individual risk of 1 E-4.

Monitoring requirements -- Community and non-transient water system monitoring based on state vulnerability assessment; vulnerable systems to be monitored quarterly for one year; repeat monitoring dependent upon detection and size of system.

Analytical methodology -- High pressure liquid chromatography (EPA 550, 550.1); gas chromatographic/mass spectrometry (EPA 525): PQL= 0.0002 mg/L.

Best available technology -- Granular activated carbon

Reference -- 55 FR 30370 (07/25/90)

EPA Contact -- Drinking Water Standards Division / OGWDW /
(202) 260-7575 / FTS 260-7575; or Safe Drinking Water Hotline / (800) 426-4791

___IV.B.3. SECONDARY MAXIMUM CONTAMINANT LEVEL (SMCL) for Drinking Water

No data available

___IV.B.4. REQUIRED MONITORING OF "UNREGULATED" CONTAMINANTS

Status -- Listed (Proposed, 1991)

Discussion -- "Unregulated" contaminants are those contaminants for which EPA establishes a monitoring requirement but which do not have an associated final MCLG, MCL, or treatment technique. EPA may regulate these contaminants in the future.

Monitoring requirement -- All systems to be monitored unless a vulnerability assessment determines the system is not vulnerable.

Analytical methodology -- Gas chromatography/mass spectrometry (EPA 525); high pressure liquid chromatography (EPA 550, 550.1).

Reference -- 56 FR 3526 (01/30/91)

EPA Contact -- Drinking Water Standards Division / OGWDW / (202) 260-7575 / FTS 260-7575; or Safe Drinking Water Hotline / (800) 426-4791

---

---

---

SMCL - NO DATA

---

FISTD- NO DATA

---

FIREV- NO DATA

---

CERC -

Value -- 1 pound (Final, 1989)

Considers technological or economic feasibility? -- NO

Discussion -- The RQ for benzo(a)pyrene is based on potential carcinogenicity (group B2). This chemical is currently under assessment for carcinogenicity and chronic toxicity and the RQ is subject to change in future rulemaking.

Reference -- 54 FR 33418 (08/14/89)

EPA Contact -- RCRA/Superfund Hotline
(800) 424-9346 / (202) 260-3000 / FTS 260-3000

---

---

SARA - NO DATA

---

RCRA -

Status -- Listed

Reference -- 52 FR 25942 (07/09/87)

EPA Contact -- RCRA/Superfund Hotline
(800)424-9346 / (202)260-3000 / FTS 260-3000

---

---

---

TSCA -

No data available

---

---

OREF - None
IREF - None
CREF - Brune, H., R.P. Deutsch-Wenzel, M. Habs, S. Ivankovic and D. Schmahl. 1981. Investigation of the tumorigenic response to benzo[a]pyrene in aqueous caffeine solution applied orally to Sprague-Dawley rats. J. Cancer Res. Clin. Oncol. 102(2): 153-157.
CREF - Clement Associates. 1990. Ingestion dose-response model to benzo(a)pyrene. EPA Control No. 68-02-4601.
CREF - IARC (International Agency for Research on Cancer). 1983. Certain Polycyclic Aromatic Hydrocarbons and Heterocyclic Compounds. Monographs on the Evaluation of Carcinogenic Risk of the Chemical to Man, Vol. 3. Lyon, France.
CREF - Knauf, L. and G. Rice. 1992. Statistical Evaluation of Several Benzo[a]pyrene Bioassays. Memorandum to R. Schoeny, U.S. EPA, Cincinnati, OH. January 2.
CREF - Neal, J. and R.H. Rigdon. 1967. Gastric tumors in mice fed benzo[a]pyrene -- A quantitative study. Tex. Rep. Biol. Med. 25(4): 553-557.
CREF - Rabstein, L.S., R.L. Peters and G.J. Spahn. 1973. Spontaneous tumors and pathologic lesions in SWR/J mice. J. Natl. Cancer Inst. 50: 751-758.
CREF - Stiteler, W. 1991. Syracuse Research Corporation, Syracuse, NY. Personal communication with R. Schoeny, U.S. EPA, Cincinnati, OH.
CREF - Thyssen, J., J. Althoff, G. Kimmerle and U. Mohr. 1981. Inhalation studies with benzo[a]pyrene in Syrian golden hamsters. J. Natl. Cancer Inst. 66: 575-577.
CREF - U.S. EPA. 1991a. Drinking Water Criteria Document for PAH. Prepared by the Office of Health and Environmental Assessment, Environmental Criteria and Assessment Office, Cincinnati, OH for the Office of Water Regulations and Standards, Washington, DC.
CREF - U.S. EPA. 1991b. Dose-Response Analysis of Ingested Benzo[a]pyrene (CAS No. 50-32-8). Human Health Assessment Group, Office of Health and Environmental Assessment, Washington, DC. EPA/600/R-92/045.
HAREF- None

# Appendix D

# Outdoor Exposure Frequencies for the Neighborhoods Near the Industrial Property

25 October 1996

Alceon Corporation
PO Box 382669
Harvard Square Station
Cambridge, MA 02238-2669

617-864-4300

# Appendix D

## Outdoor Exposure Frequencies for the Neighborhoods Near the Industrial Property

We all know that weather strongly affects the nature and duration of outdoor activities. Since weather data are routinely recorded at both O'Hare and Midway Airports, we use the weather records from Midway Airport -- the closer of the two airports in Chicago -- as a surrogate for activity patterns. Table 8 in the main report summarizes the average daily temperatures recorded at Midway Airport from 1961 - 1990 (US Department of Commerce, 1992). More specifically, the top two rows of data in Table 8 show the number of days per year that have average temperatures (denoted T) at or above the stated temperature and -- by difference -- the number of days per year that have average temperatures below the stated value. For example, in a typical year at Midway Airport, there are 196 day/yr with $T \geq 50$ degF and 169 day/yr with $T < 50$ degF.

For children, for example, we assume that T strongly influences the number of days in a year on which a child behaves in such a way as to ingest incidentally soils outdoors. In particular, we assume the following information in Table 8.

- For 83 days when $T < 32$ degF, we assume that no child incidentally ingests soil outdoors. On such cold days, the soils outdoors are frozen and/or covered with snow and ice. Children may play outside on such days, but they cannot ingest the frozen soils.
- For 119 days when $T < 40$ degF, we assume that 5 percent of children incidentally ingest some surface soils outdoors.
- For 169 days when $T < 50$ degF, we assume that 20 percent of children incidentally ingest some surface soils outdoors.
- For 222 days when $T < 60$ degF, we assume that 70 percent of children incidentally ingest some surface soils outdoors.
- For 86 days when $T \geq 70$ degF, we assume that 100 percent of children incidentally ingest some surface soils outdoors.

From this information and the data in Table 8, we develop a probability distribution for the number of days that a child plays outdoors in a typical year as follows.

First, we use concepts and methods routinely used to analyze reliability data (Cox & Oakes, 1984; Crowder et al, 1991; Lee, 1992). To start, we define the condition of

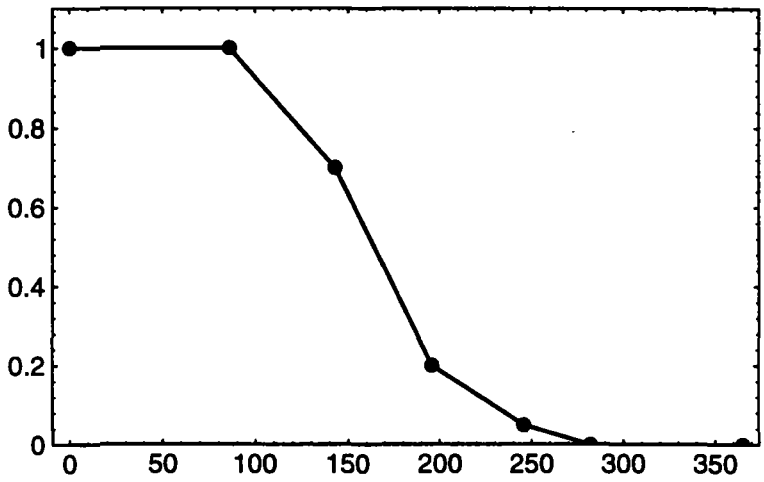
"success" as "playing outdoors in warm weather", and we define the event of "failure" as "the end to playing outdoors due to cold weather."

With these routine definitions, we develop the probability density function (PDF) for the "time to failure," denoted t and measured in days, in three steps:

- First, we develop S(t), the complementary cumulative distribution function (CCDF) for the t, the "time to failure," by direct interpretation of the behavior and the weather data. This curve declines monotonically from S(0) = 1 to S(365) = 0.
- Second, we derive F(t), the cumulative distribution function (CDF), for the "time to failure," by subtraction: F(t) = 1 - S(t). This curve rises monotonically from F(0) = 0 to F(365) = 1.
- Third, we derive f(t), the probability density function (PDF) for the "time to failure," by differentiating F(t) with respect to time: $f(t) = \frac{d}{dt} F(t)$. The area under this curve equals 1.

Thus f(t) is the PDF for the number of days in a typical year that a child plays outdoors.

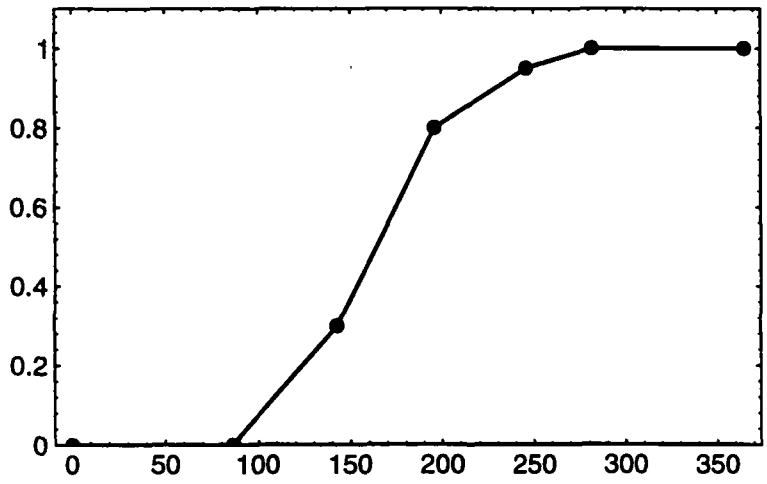
For children, the graph below shows S(t), the CCDF for the time to failure, i.e., the complementary cumulative probability distribution for the number of days in a typical year that a child plays outdoors in the neighborhoods near the industrial property.



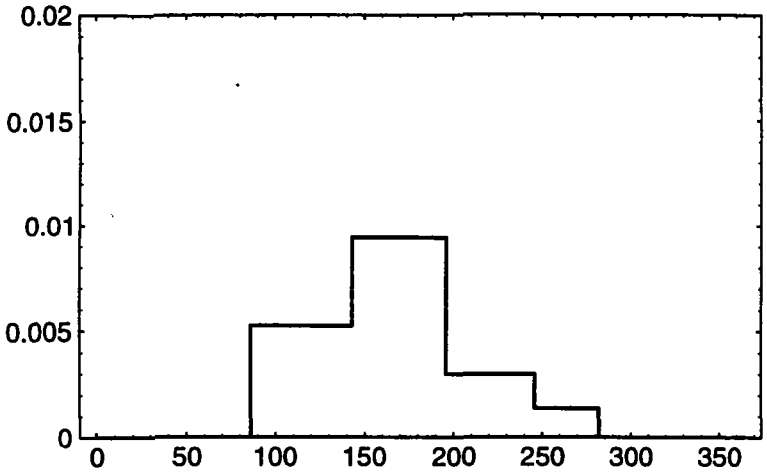
The graph of S(t) for children above plots this information:

- On warm days, when T equals or exceeds 70 degF, all children play outdoors. This corresponds to a plotted point of (t, S(t)) = (86, 1.00).
- On cool days, when T declines to 60 degF, 70 percent of the children continue outdoor activities that culminate in soil ingestion. This corresponds to plotted point of (t, S(t)) = (143, 0.70).
- On cooler days, when T declines to 50 degF, only 20 percent of the children continue outdoor activities that culminate in soil ingestion. This corresponds to a plotted point of (t, S(t)) = (196, 0.20).
- On cold days, when T declines to 40 degF, only 5 percent of the children continue outdoor activities that culminate in soil ingestion. This corresponds to a plotted point of (t, S(t)) = (246, 0.05).
- On freezing days, when T declines below 32 degF, no children continue outdoor activities that culminate in soil ingestion. The children go outdoors, but they cannot ingest frozen soils. This corresponds to a plotted point of (t, S(t)) = (282, 0.00).

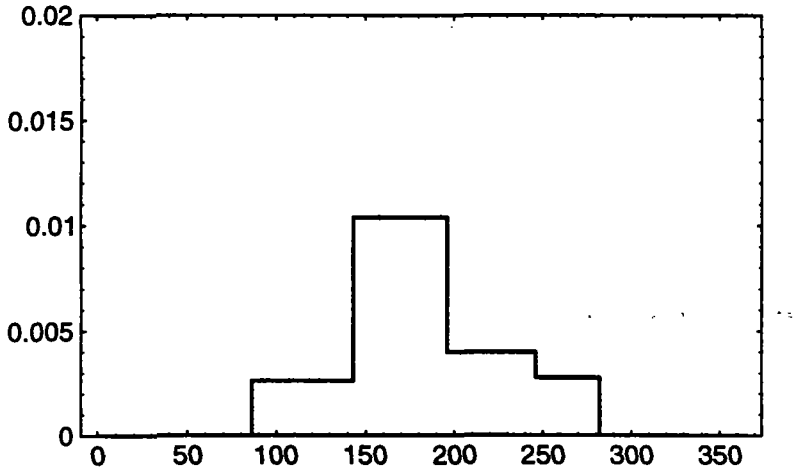
For children, the graph below shows F(t) = 1 - S(t), the CDF for the time to failure, i.e., the cumulative probability distribution for the number of days in a typical year that a child plays outdoors in the neighborhoods near the industrial property.



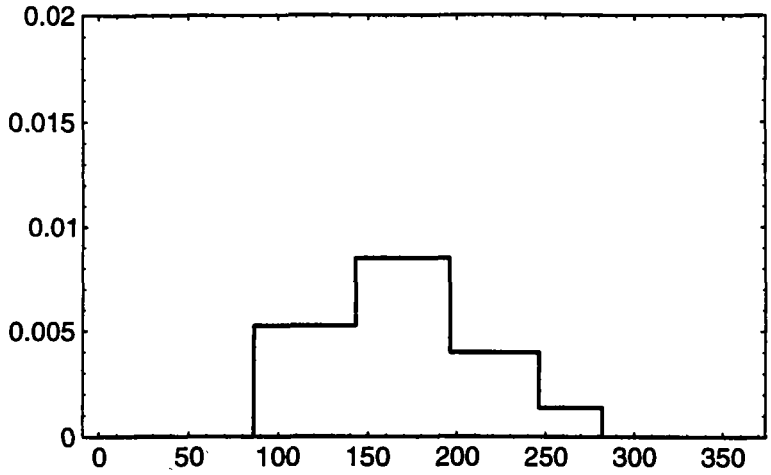

For children, the graph below plots $f(t) = \frac{d}{dt} F(t)$, the PDF for the time to failure, i.e., the probability density for the number of days in a typical year that a child plays outdoors in the neighborhoods near the industrial property. This random variable has a median equal to 164.2 days and a mean equal to 165.5 days.




Using the information in Table 8 for teens, the graph below plots $f(t) = \frac{d}{dt} F(t)$, the PDF for the time to failure, i.e., the probability density for the number of days in a typical year that a teen plays outdoors in the neighborhoods near the industrial property. This random variable has a median equal to 176.7 days and a mean equal to 181.0 days




Using the information in Table 8 for adults, the graph below plots $f(t) = \frac{d}{dt} F(t)$, the PDF for the time to failure, i.e., the probability density for the number of days in a typical year that an adult plays outdoors in the neighborhoods near the industrial property. This random variable has a median equal to 166.6 days and a mean equal to 168.0 days.




## References

Cox & Oakes, 1984
Cox, D.R. and D. Oakes, 1984, Analysis of Survival Data, Chapman & Hall, London, UK

Crowder et al, 1991
Crowder, M.J., A.C. Kimber, R.L. Smith, and T.J. Sweeting, 1991, Statistical Analysis of Reliability Data, Chapman & Hall, London, UK

Lee, 1992
Lee, E.T., 1992, Statistical Methods for Survival Data Analysis, Second Edition, John Wiley & Sons, New York, NY

# Appendix E

# Exposure Durations for the Neighborhoods Near the Industrial Property

25 October 1996

Alceon Corporation
PO Box 382669
Harvard Square Station
Cambridge, MA 02238-2669

617-864-4300

# Appendix E

## Exposure Durations for the Neighborhoods Near the Industrial Property

As employees of the US Environmental Protection Agency, Israeli and Nelson (1992) estimated distributions of the residence times for different groups of US households based on data published by the Bureau of the Census. Israeli and Nelson report that the distribution for total residence time is essentially an Exponential distribution with a different mean value for each different housing group. An Exponential distribution is completely characterized by its mean value and is highly skewed (i.e., far from symmetric), with a long tail to the right.

In this report, we accept 49 percent as the fraction of owner-occupied houses in the neighborhoods near the industrial property (Ecology & Environment, 1995, Letter). From this, we calculate 51 percent as the fraction of non-owner-occupied houses in the same neighborhoods.

On a neighborhood- and site-specific basis, we estimate the 90th percentile of the mixed population distribution using information in the column titled "Average Total Residence Time, T (years)" of Table IV of Israeli and Nelson (1992).

First, we draw 4,900 realizations from this distribution:

Owners ~ Exponential ( 1 / 11.36 yr )

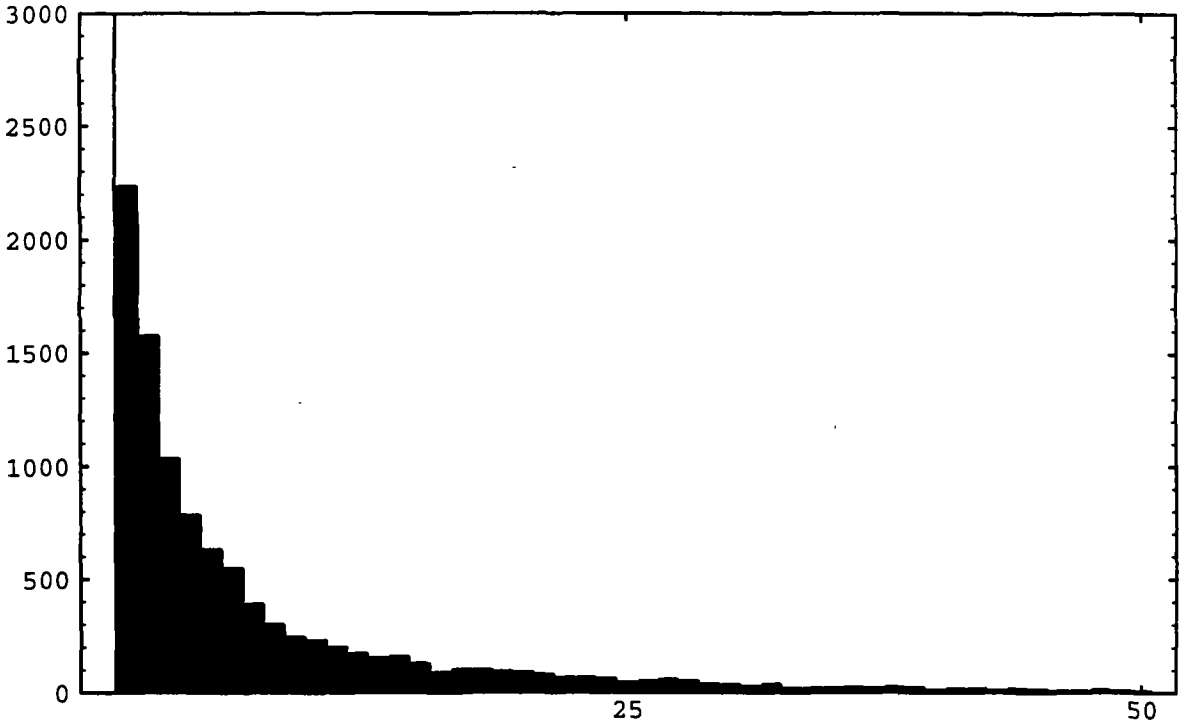
and 5,100 realizations from this distribution:

Renters ~ Exponential ( 1 / 2.35 yr ).

Thus, we have simulated a mixed population of 49 percent owners and 51 percent renters.

This mixed distribution has these summary statistics: arithmetic mean = 6.64 yr, 10th percentile = 0.42 yr; 25th percentile = 1.15 yr; 50th percentile = 3.20 yr; 75th percentile = 8.11 yr; and 90th percentile = 18.00 yr.

This graph shows the histogram for this mixed distribution.



# Appendix F

# Relationship Between Concentrations of Conservative Tracers in Indoor Dust and Outdoor Soil

25 October 1996

Alceon Corporation
PO Box 382669
Harvard Square Station
Cambridge, MA 02238-2669

617-864-4300

Accepted by the Journal of Soil Contamination

# A Parametric Distribution for the Fraction of Outdoor Soil in Indoor Dust

Philip R. Trowbridge
David E. Burmaster
Alceon Corporation
Post Office Box 382669
Harvard Square Station
Cambridge, MA 02238-2669
617-864-4300
deb@Alceon.com

## Abstract

For a chemical which does not have a source inside a house, the ratio of its dust concentration indoors to its soil concentration outdoors is equal to the fraction of house dust which is composed of soil. To estimate the fraction of soil in house dust, we compiled ratios of the concentrations of a chemical in dust and soil from the scientific literature. We find that a LogNormal distribution fits the data extremely well. This distribution is suitable for use in public health risk assessments for single-family homes in temperate climates.

## Introduction

Ingestion of indoor dust is a significant exposure pathway for children in residential settings (Calabrese and Stanek, 1992, Dust; Stanek and Calabrese, 1992; Chuang et al., 1995; Fergusson and Kim, 1991). In one study, Stanek and Calabrese (1992) demonstrated that almost 50 percent of the soil ingested by children came from ingestion of soil in indoor dust. Measurements of contaminant concentrations in dust are difficult to perform and uncommon in human health risk assessment studies. Consequently, there is a need for a method to estimate the concentration of a chemical in dust from the concentration of that chemical in the soil outside the house.

The composition of indoor dust differs strongly from the composition of the soil outside a house. Only a fraction of dust is composed of soil which has been carried into the house (e.g., on shoes). The remainder consists of dust particles derived from material inside the house such as lint from carpets and clothes, human hair and skin, pet hair and skin, household plant material, pieces of paper, paint chips, wood chips from furniture, pieces of insulation, flakes of construction materials, bacteria, viruses, allergens and insects (e.g., dust mites) (Thatcher and Layton, 1995). Some indoor dust is also derived from particles carried through open windows by the wind.

Because indoor dust is a mixture of particles generated inside the house ("particles") and soil carried into the house on clothing, the concentration of a chemical in indoor dust must fall between its concentration in these two media. If the concentration of the chemical in the particles is negligible compared with its concentration in soil, the dust concentration of the chemical can be predicted from the physical dilution of the soil by particles. We define chemicals for which the particle concentration is negligible compared to the soil concentration as "conservative tracer chemicals".

Assuming that the soil carried into the house has the same chemical and physical properties as the outdoor soil, the ratio of the dust concentration to the soil concentration for conservative tracer chemicals is equal to the fraction of the household dust which consists of soil. For convenience, we define this ratio as the "transfer coefficient" (TC) of an chemical (Eqn 1).

$$TC = \frac{C_{dust}}{C_{soil}} \qquad \text{Eqn 1}$$

where $C_{dust}$ is the dust concentration of the element with units of mg/kg and $C_{soil}$ is the soil concentration of the element with units of mg/kg. The TC is dimensionless. For conservative tracer chemicals, the maximum value for the TC, 1, represents dust

composed entirely of soil. The minimum value for the TC, 0, can only occur for dust which does not contain any soil. The TC for an element can only exceed a value of 1 if there are sources of that element from material inside the house.

We searched the scientific literature for studies in which the concentration of a chemical in both the dust and the soil was measured. To estimate the fraction of soil in house dust, we calculated the TC for all data pairs ($C_{dust}$, $C_{soil}$) which met the criteria for being conservative tracer chemicals as discussed below. We model the variability in the value of the TC by representing this ratio as a distribution.

Selection of Data for Conservative Tracers

We compiled data pairs ($C_{dust}$, $C_{soil}$) for rare earth elements (Binder et al., 1986; Bowen, 1979, Calabrese et al., 1989; Calabrese and Stanek, 1992, Dust; Calabrese and Stanek, 1992, Pica; Davis et al., 1990; Fergusson et al., 1986; Fergusson and Kim, 1991), heavy metals (Hartwell et al., 1983; Hawley, 1985; Lioy et al., 1992; Stern, 1994), several pesticides (Camann and Lewis, 1993; Simcox et al., 1995), and some organic compounds (Chuang et al., 1995). All the studies were conducted for single-family homes in temperate climates.

For each data pair, we determined whether it was a conservative tracer chemical based on three criteria. First, the data pair must be for one of the soil-derived elements proposed by Fergusson et al. (1986): halfnium (Hf), thorium (Th), scandium (Sc), samarium (Sm), cerium (Ce), lanthanum (La), manganese (Mn), sodium (Na), potassium (K), vanadium (V), aluminum (Al) and iron (Fe). Fergusson et al. (1986) showed that these elements do not have any sources or sinks within houses other than soil. Other chemicals, such as lead (Pb), arsenic (As), chromium (Cr), polycyclic aromatic hydrocarbons (PAHs), and some pesticides, can have higher dust concentrations than soil concentrations due to sources inside the house (e.g., lead

paint, tobacco smoke) (Hartwell et al., 1983; Lioy et al., 1992; Chuang et al., 1995; Simcox et al., 1995). Therefore, these compounds are not expected to behave as conservative tracer chemicals. Second, the mean dust and soil concentrations for the element must be significantly different at the $p < 0.05$ confidence level. Third, the dust concentration must be less than the soil concentration for the element. In theory, the TC can reach a maximum value of one for pure soil. In reality, it would be impossible for a conservative tracer chemical to have a TC value of one because of the large fraction of organic material which is always present in indoor dust (Rothenberg et al., 1989).

Table 1 shows the data pairs that meet these criteria and their associated TCs. Most of the values in Table 1 are mean values except those from Fergusson and Kim (1991) which are median values. We think the insights gained from including the median values outweigh the uncertainties introduced by their inclusion. It is not possible to test the second criteria for the data from Fergusson and Kim (1991) because only the median values for dust and soil concentrations are shown in this article. The values in Table 1 rely on studies with $11 \leq N \leq 101$ data points.

Estimation of the Transfer Coefficient Distribution

By comparing the histogram of the TC values in Table 1 to Normal, Beta, and LogNormal distributions, we conclude that the LogNormal distribution represents the variability in the TC exceedingly well. In Figure 1, ln[TC] has been plotted versus zscore. The solid line corresponds to a perfect LogNormal distribution while the points are the data from Table 1. All the points are clustered close to this line ($r^2 = 0.9729$) which indicates that the data are well described by a LogNormal distribution. Ott (1995) notes that dilution processes tend to produce concentration distributions which are LogNormal in character. The fact that the values of TC are distributed LogNormally is consistent with Ott's observation because the variability in the TC results from the dilution of soil by particles.

We parameterize the LogNormal distribution as Eqn 2.

$$\underline{TC} \sim \exp[\text{ Normal}(\mu, \sigma) ] \qquad \text{Eqn 2}$$

which is identical to

$$\ln[\underline{TC}] \sim \text{Normal}(\mu, \sigma) \qquad \text{Eqn 3}$$

where $\underline{TC}$ is a LogNormal random variable, $\mu$ is the arithmetic mean of the Normal random variable $\ln[\underline{TC}]$, and $\sigma$ is the arithmetic standard deviation of Normal random variable $\ln[\underline{TC}]$. By fitting a line to the points in Figure 1 using Mathematica™, we estimate that $\hat{\mu} \pm se = -0.8767 \pm 0.0122$ and $\hat{\sigma} \pm se = 0.3663 \pm 0.0125$. Figure 2 shows the LogNormal probability density function (PDF) described by these parameters. This figure illustrates the properties of $\underline{TC}$. First, the minimum value for $\underline{TC}$ is 0. Second, the mode of $\underline{TC}$ is 0.3639. Third, the median of $\underline{TC}$ is 0.4162. Fourth, the arithmetic mean and arithmetic standard deviation of $\underline{TC}$ are 0.4450 and 0.1687, respectively. Figure 3 shows the cumulative distribution function (CDF) for this fitted distribution with the data from Table 1 superimposed on the graph. The PDF and CDF are alternative ways to represent the same information.

Even though a LogNormal distribution is defined from 0 to infinity, the probability of having a TC value greater than 1 is less than 1 percent given these fitted parameters. In practice, we recommend truncating the distribution of $\underline{TC}$ at a maximum of 1 because TC values $\geq 1$ are only possible for non-conservative tracer chemicals. Therefore, truncating the distribution at a maximum of 1 will change the distribution by only a negligible amount.

## Discussion

The fraction of soil in house dust is a random variable which closely follows a LogNormal distribution with an arithmetic mean of 0.445 and an arithmetic standard deviation of 0.1687. The arithmetic mean and arithmetic standard deviation ln[TC] are -0.8767 and 0.3663, respectively. This parametric distribution is suitable for use in human health risk assessments for single family homes in temperate climates. These findings are consistent with the conclusions of Calabrese and Stanek (1992, Dust) that the mean fraction of indoor dust originally derived from soil is 0.313. Therefore, for a conservative tracer chemical, its concentration in house dust is expected to be less than half its concentration in the soil outside the home.

In this paper, we compare the bulk chemical compositions of soil and indoor dust because these were the only properties which were measured in the original studies. If the data were available, it would be more accurate to compare the chemical compositions of the two media for each particle size. Estimating the TC in this manner would eliminate the potentially confounding effect that some particle sizes may be preferentially transported into houses relative to other sizes.

## Acknowledgements

We thank Edward J. Calabrese for his assistance and advice on this project.
Alceon Corporation and AlliedSignal funded this research.

## Trademarks

Mathematica™ is a registered trademark of Wolfram Research, Inc. in Champaign, IL.

## References

Binder et al., 1986

Binder, S., Sokal, D., Maughan, D., 1986, Estimating the Amount of Soil Ingested by Young Children Through Tracer Elements, Archives of Environmental Health, Volume 41, p. 341-345.

Bowen, 1979

Bowen, H.J.M, 1979, Environmental Chemistry of the Elements, Academic Press, London, UK, 333 p.

Calabrese et al., 1989

Calabrese, E.J., Pastides, H., Barnes, R., Edwards, C., Kosteki, P.T., Stanek, E.J., Veneman, P., and Gilbert, C.E., 1989, How Much Soil Do Young Children Ingest: An Epidemiologic Study, In E.J. Calabrese and P.T. Kostecki, eds., Petroleum Contaminated Soils, Lewis Publishing, Chelsea, MI, p. 363-397.

Calabrese and Stanek, 1992, Dust

Calabrese, E.J., and Stanek, E.J., 1992, What Proportion of Household Dust Is Derived from Outdoor Soil? Journal of Soil Contamination, Volume 1, Number 3, p. 253-263.

Calabrese and Stanek, 1992, Pica

Calabrese, E.J., and Stanek, E.S., 1992, Distinguishing Outdoor Soil Ingestion from Indoor Dust Ingestion in a Soil Pica Child, Regulatory Toxicology and Pharmacology, Volume 15, p. 83-85.

Camann and Lewis, 1993

Camann, D.E., and Lewis, R.G., 1993, Trapping of Particle-Associated Pesticides in Indoor Air by Polyurethane Foam and Exploration of Soil Track In as a Pesticide Source, Proceedings of the

1993 Annual Conference of the Air and Waste Management Association on Measurements of Toxic and Related Air Pollutants, p. 621-626.

Chuang et al., 1991

Chuang, J.C., Gregory, A.M., Kuhlman, M.R., and Wilson, N.K, 1991, Polycyclic Aromatic Hydrocarbons and Their Derivatives in Indoor and Outdoor Air in an Eight-Home Study, Atmospheric Environment, Volume 25B, Number 3, p. 369-380.

Chuang et al., 1995

Chuang, J.C., Callahan, P.J., Menton, R.G., Gordon, S.M., Lewis, R.G. and Wilson, N.K., 1995, Monitoring Methods for Polycyclic Aromatic Hydrocarbons and Their Distribution in House Dust and Track-In Soil, Environmental Science and Technology, Volume 29, Number 2, p. 494-500.

Davis et al., 1990

Davis, S., Waller, P., Buschbom, R., Ballou, J., White, P.,1990, Quantitative Estimates of Soil Ingestion in Normal Children Between the Ages of 2 and 7 Years: Population Based Estimates Using Aluminum, Silicon, and Titanium as Soil Tracer Elements, Archives of Environmental Health, Volume 45, Number 2, p. 112-122.

Fergusson et al., 1986

Fergusson, J.E., Forbes, E.A., Schroeder, R.J., and Ryan, D.E.,1986, The Elemental Composition and Sources of House Dust and Street Dust, Science of the Total Environment, Volume 50, p. 217-221.

Fergusson and Kim, 1991

Fergusson, J.E., and Kim, N.D., 1991, Trace Elements in Street and House Dusts: Sources and Speciation, Science of the Total Environment, Volume 100, p.125-150.

Hartwell et al., 1983

Hartwell, T.D., Handy, R.W., Harris, B.S., Williams, S.R., and Gehlbach, S.H., 1983, Heavy Metal Exposure in Populations Living Around Zinc and Copper Smelters, Archives of Environmental Health, Volume 38, Number 5, p. 284-295.

Hawley, 1985

Hawley, J.K., 1985, Assessment of Health Risk from Exposure to Contaminated Soil, Risk Analysis, Volume 5, Number 4, p. 289-302.

Lioy et al., 1992

Lioy, P.J., Freeman, N.C.G., Wainman, T., Stern, A.H., Boesch, R., Howell, T., and Shupack, S.I., 1992, Microenvironmental Analysis of Residential Exposure to Chromium Laden Wastes in and Around New Jersey Homes, Risk Analysis, Volume 12, Number 2, pp. 287-299.

Ott, 1995

Ott, W.R., 1995, Environmental Statistics and Data Analysis, Lewis Publishers, Boca Raton, FL, 313 p.

Rothenberg et al., 1989

Rothenberg, S.J., Nagy, P.A., Pickrell, J.A., and Hobbs, C.H., 1989, Surface Area, Adsorption, and Desorption Studies on Indoor Dust Samples, American Industrial Hygiene Association Journal, Volume 50, Number 1, p. 15-23.

Simcox et al., 1995

Simcox, N.J., Fenske, R.A., Wolz, S.A., Lee, I.C., and Kalman, D.A., 1995, Pesticides in Household Dust and Soil: Exposure Pathways for Children and Agricultural Families, Environmental Health Perspectives, Volume 103, Number 12, p. 1126-1134.

Stanek and Calabrese, 1992

Stanek, E.J., and Calabrese, E.J., 1992, Soil Ingestion in Children: Outdoor Soil or Indoor Dust? Journal of Soil Contamination, Volume 1, Number 1, p. 1-28.

Stern, 1994

Stern, A.H., 1994, Derivation of a Target Level of Lead in Soil at Residential Sites Corresponding to a De Minimus Contribution to Blood Lead Concentration, Risk Analysis, Volume 14, Number 6, p. 1049-1056.

Thatcher and Layton, 1995

Thatcher, T.L., and Layton, D.W., 1995, Deposition, Resuspension, and Penetration of Particles Within a Residence, Atmospheric Environment, Volume 29, Number 13, p. 1487-1497.

**Table 1: Estimated Transfer Coefficients**

| Element | Dust Concentration (mg/kg) | Soil Concentration (mg/kg) | Transfer Coefficent | Source |
|---|---|---|---|---|
| Al | 19,000.0 | 66,000.0 | 0.2879 | 3 |
| Al | 25,000.0 | 71,000.0 | 0.3521 | 2, 5 |
| Al | 23,900.0 | 55,600.0 | 0.4299 | 1 |
| Al | 33,600.0 | 66,600.0 | 0.5045 | 4 |
| Al | 47,200.0 | 54,000.0 | 0.8741 | 4 |
| Ce | 23.6 | 52.3 | 0.4512 | 1 |
| Ce | 25.0 | 50.0 | 0.5000 | 2, 5 |
| Fe | 10,000.0 | 40,000.0 | 0.2500 | 2, 5 |
| Fe | 10,200.0 | 20,000.0 | 0.5100 | 1 |
| Hf | 2.0 | 6.0 | 0.3333 | 2, 5 |
| Hf | 2.1 | 4.1 | 0.5146 | 1 |
| K | 12,600.0 | 25,000.0 | 0.5040 | 1 |
| K | 13,000.0 | 14,000.0 | 0.9286 | 2, 5 |
| La | 10.0 | 40.0 | 0.2500 | 2, 5 |
| La | 11.9 | 27.4 | 0.4343 | 1 |
| Mn | 200.0 | 1,000.0 | 0.2000 | 2, 5 |
| Mn | 207.0 | 325.0 | 0.6369 | 1 |
| Na | 11,800.0 | 18,600.0 | 0.6344 | 1 |
| Sc | 2.9 | 6.8 | 0.4240 | 1 |
| Sc | 3.0 | 7.0 | 0.4286 | 2, 5 |
| Sm | 1.2 | 4.5 | 0.2667 | 2, 5 |
| Sm | 1.2 | 3.9 | 0.3128 | 1 |
| Th | 3.0 | 9.0 | 0.3333 | 2, 5 |
| Th | 3.4 | 8.2 | 0.4172 | 1 |
| V | 30.0 | 90.0 | 0.3333 | 2, 5 |
| V | 30.4 | 66.0 | 0.4606 | 1 |

Sources
1. Fergusson et al., 1986
2. Fergusson and Kim, 1991
3. Davis et al., 1990
4. Calabrese et al., 1989
5. Bowen, 1979

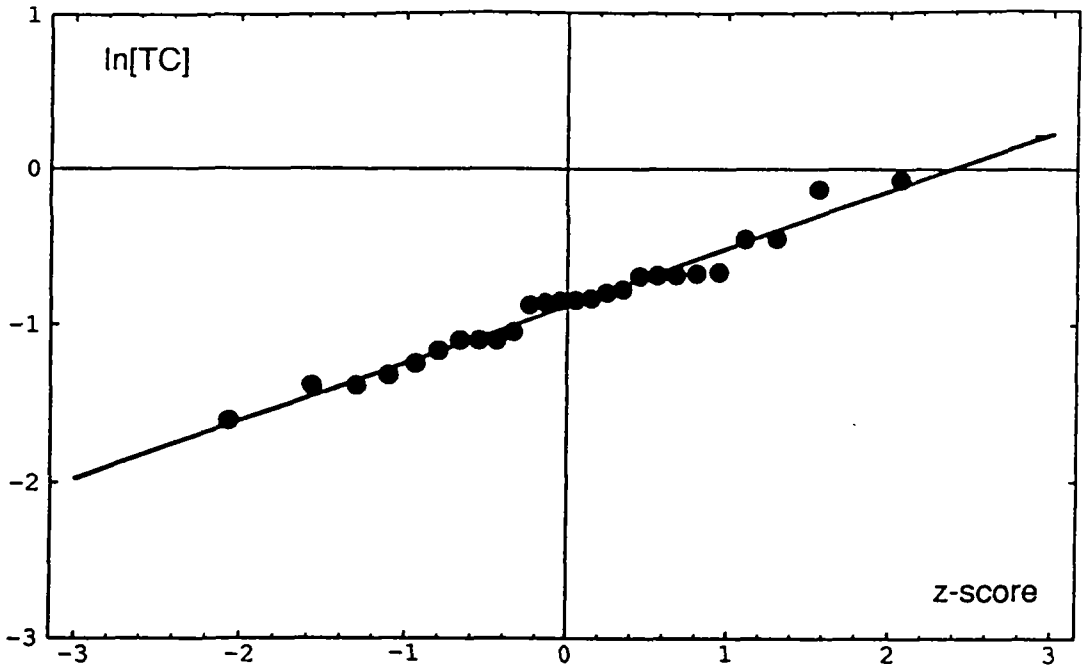



Figure 1

LogNormal Probability Plot for TC

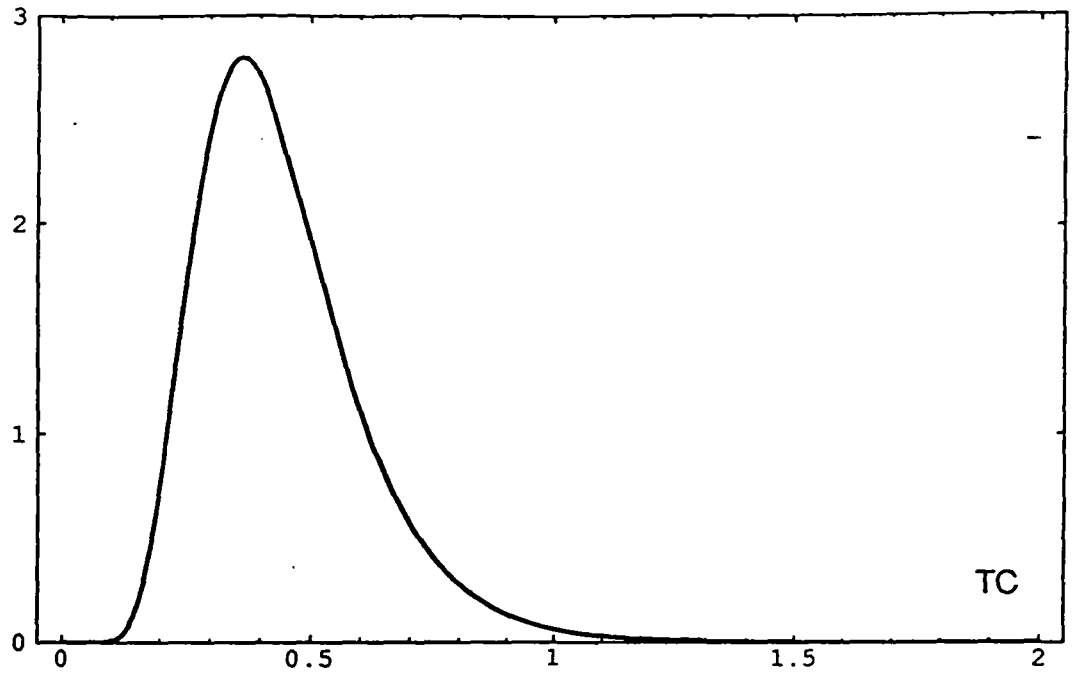



Figure 2

Probability Density Function for TC

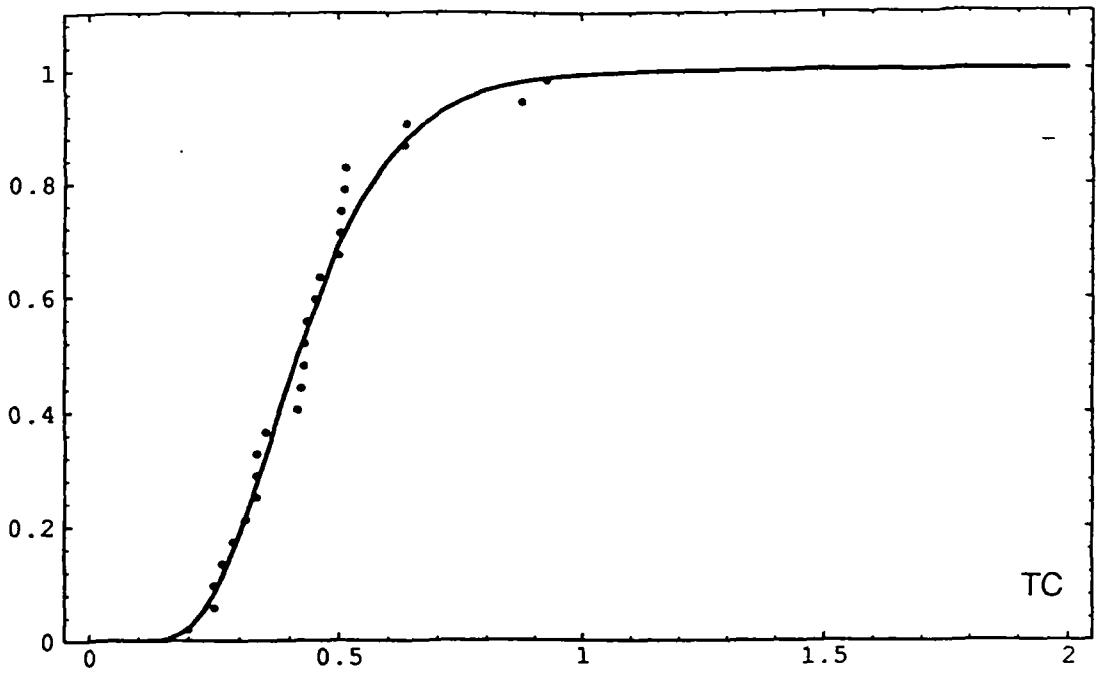



Figure 3

Cumulative Distribution Function for TC
with data points superimposed

# Appendix G

# Absorption Adjustment Factors for Polycyclic Aromatic Hydrocarbons (PAHs)

1 July 1996

Ogden Environmental and Energy Services
Westford, MA 01886

# ABSORPTION ADJUSTMENT FACTOR FOR POLYCYCLIC AROMATIC HYDROCARBONS (PAHS)

Brian Magee, Ph.D.
Ogden Environmental and Energy Services
239 Littleton Road, Suite 7C
Westford, Massachusetts 01886
508-692-9090 (voice)

June 4, 1996

## INTRODUCTION

To estimate the potential risk to human health that may be posed by the presence of chemical compounds in soil or other environmental media, it is first necessary to estimate the human exposure dose of each compound. The exposure dose is similar to the administered dose or applied dose of a laboratory experiment. The exposure dose is then combined with an estimate of the toxicity of the compound to produce an estimate of risk posed to human health.

The estimate of toxicity of a compound, termed the dose-response value, can be derived from human epidemiological data, but it is most often derived from experiments with laboratory animals. The dose-response value can be calculated based on the administered dose of the compound (similar to the human exposure dose) or, when data are available, based on the absorbed dose, or internal dose, of the compound.

In animals, as in humans, the administered dose of a compound is not necessarily completely absorbed. Moreover, differences in absorption may exist between laboratory animals and humans, as well as between different media and routes of exposure. Therefore, it is often inappropriate to directly apply a dose-response value to the human exposure dose. In many cases, a correction factor in the calculation of risk is needed to account for differences between absorption in the dose-response study and absorption likely to occur upon human exposure to a compound. Without such a correction, the estimate of human health risk could be over- or under-estimated.

This correction factor is defined here as the absorption adjustment factor, or AAF. The AAF is used to adjust the human exposure dose so that it is expressed in the same terms as the doses used to generate the dose-response curve in the dose-response study. The AAF is the ratio between the estimated absorption factor for the specific medium and route of exposure, and the known or estimated absorption factor for the laboratory study from which the dose-response value was derived.

In some cases, AAFs can be derived from data within a single experiment if an appropriate measure of absorption is compared between different routes of administration and/or sample matrices. In other cases, a single experiment may quantitate total fractional absorption for only

one matrix and route of exposure. AAFs can be derived from such experiments if coupled with data from other experiments that quantitate the absorption from the route and matrix used in the dose-response study. In this case, the AAF is derived using the following equation:

AAF = (fraction absorbed from the environmental exposure)/
(fraction absorbed in the dose-response study).

The use of an AAF allows the risk assessor to make appropriate adjustments if the efficiency of absorption between environmental exposures and experimental exposures is known or expected to differ because of physiological effects and/or matrix or vehicle effects. Absorption adjustment factors can be less than one or greater than one. If the absorption from the site-specific exposure is the same as absorption in the laboratory study, then the AAF is 1.0. An AAF of 1.0 does not indicate that absorption is 100%. It indicates that absorption is known or estimated to be the same as that in the dose-response study.

EPA explicitly discusses the appropriateness of using absorption/bioavailability factors in the Guidelines for Exposure Assessment (EPA, 1992a). For instance, EPA states:

> The applied dose, or the amount that reaches exchange boundaries of the skin, lung, or gastrointestinal tract, may often be less than the potential dose if the material is only partly bioavailable. Where data on bioavailability are known, adjustments to the potential dose to convert it to applied dose and internal dose may be made.
>
> This may be done by adding a bioavailability factor (range: 0 to 1) to the dose equation. The bioavailability factor would then take into account the ability of the chemical to be extracted from the matrix, absorption through the exchange boundary, and any other losses between ingestion and contact with lung or gastrointestinal tract.

The Guidelines for Exposure Assessment discuss the issues of absorption and bioavailability throughout the document, indicating EPA's current understanding that the inclusion of properly documented absorption adjustment factors is a scientifically appropriate and important aspect of the risk assessment process. The Absorption Adjustment Factors derived here take into account matrix-specific bioavailability as well as knowledge of PAH pharmacokinetics. These AAFs

should be used in the calculation of the Average Daily Doses (ADD) that are necessary to quantitatively estimate potential risk to human health.

In this paper, the route of exposure and the experimental matrix (diet, drinking water, corn oil gavage, etc.) used in the experimental studies from which the relevant dose-response value was derived are summarized for the polycyclic aromatic hydrocarbons (PAHs). In addition, the scientific literature on the absorption and bioavailability of PAHs has been reviewed for the relevant routes of exposure and matrices. Based on these data, scientifically defensible oral-soil and dermal-soil AAFs have been derived. The information and methods used to derive these AAFs are described below.

Although it is possible in theory, absorption experiments in humans that are suitable for AAF derivation have not been executed. Thus, AAFs are derived from animal studies. Because AAFs can be derived from multiple scientific studies using different animals species and strains and different experimental conditions, there is scientific uncertainty concerning the true AAF for the human exposure situation. This uncertainty can be incorporated into the risk assessment process by deriving distributions for the relevant AAFs. Accordingly, oral-soil and dermal-soil AAFs for PAHs are derived here both as point estimates for deterministic risk assessments and as distributions for probabilistic risk assessments.

## ABSORPTION FROM THE DOSE-RESPONSE STUDIES

Potentially carcinogenic PAH are routinely evaluated using the comparative potency approach described in EPA (1993). With this approach, all potentially carcinogenic PAH are assessed in terms of their benzo(a)pyrene toxic equivalent concentrations, and EPA's cancer slope factor for benzo(a)pyrene is used. .

### Derivation of Cancer Slope Factor for Benzo(a)pyrene

The risk assessment of potentially carcinogenic PAHs is performed using the oral cancer slope factor (CSF) for benzo(a)pyrene (B(a)P). The oral CSF for B(a)P (7.3 $(mg/kg\text{-}day)^{-1}$) is the geometric mean of four slope factors derived from two rodent feeding studies: Neal and Rigdon (1967) and Brune *et al.* (1981). In the first study, CFW mice were dosed with B(a)P in their

laboratory chow (diet). The diet was prepared by dissolving benzo(a)pyrene in benzene, mixing with wheat flour, evaporating the benzene and mixing the flour-benzo(a)pyrene mixture with laboratory chow pellets. In the second, Sprague Dawley rats were also dosed with B(a)P in their laboratory chow (diet).

**Gastrointestinal Absorption in Dose-Response Study**

Absorption of B(a)P from food has been shown to be high in both humans and rodents by several researchers. Many articles on absorption were reviewed. However, studies that used inappropriate scientific methods were rejected for AAF derivation. For instance, studies that measured total radiolabel in the feces do not yield useful absorption information, because B(a)P metabolites are known to be excreted into bile (see, for instance, Chipman *et al.*, 1981a, 1981b; Bowes and Renwick, 1986).

As an example, data are presented in a paper by Chang (1943) on fecal excretion of benzo(a)pyrene and other PAH. This paper cannot be used to estimate gastrointestinal absorption of PAH, because the gravimetric method used is nonspecific and does not distinguish between unchanged PAH and PAH metabolites. A paper by Flesher and Syndor (1960) is also deficient for AAF derivation, because total tritium is measured in feces after oral dosing of rats with $^{3}$H-3-methylcholanthrene. This method also does not distinguish between unabsorbed PAH and absorbed and metabolized PAH excreted into the bile and feces.

Other studies are not useful because they only define a small fraction of a PAH's total disposition. For instance, in a study by Rees *et al.* (1971), benzo(a)pyrene was given to rats by stomach tube and the PAH was measured in the lymphatic duct. While the presence of B(a)P in the lymph indicates that absorption occurred, the experiment is not quantitative. Similarly, Foth *et al.* (1988) measured benzo(a)pyrene absorption in the rat after a continuous infusion into the duodenum by measuring B(a)P in the atrial blood and bile. In this case, the conditions of the experiment are unnatural, and the experiment does not account for a total mass balance of B(a)P. Other studies were rejected for similar reasons. The following principal studies are those in which useful absorption information can be gleaned.

*Hecht et al. (1979)*

Hecht and coworkers (Hecht *et al.*, 1979) fed B(a)P to both humans and F-344 rats and measured the unchanged B(a)P in the feces to obtain an estimate of the amount of the compound absorbed. Because unchanged B(a)P in the feces can be due to absorbed material that is excreted unchanged in the bile, these studies reveal the minimum amount of B(a)P that was absorbed. It is known, however, that B(a)P is extensively metabolized. Thus, these estimates of absorption are valid for AAF derivation.

For rats, at least 87% of the B(a)P was absorbed from a low single dose in peanut oil (0.037 mg/kg). Minimum absorption from medium and high doses (0.37 mg/kg and 3.7 mg/kg) were 92.2% and 94.4%. The mean absorption of B(a)P in peanut oil in rats was 91.2% (n=30). This value was used in AAF derivation.

When rats were fed a single dose of charcoal-broiled hamburger containing B(a)P (0.002 mg/kg body weight), at least 89% was absorbed (n=10). In humans, a high percentage of B(a)P present in charcoal-broiled meat was also absorbed (0.0001 mg/kg body weight, assuming 70 kg), because no unchanged B(a)P was detected in the feces. Assuming that B(a)P was present in feces at 1/2 the detection limit, the minimal absorption is 98.8% (n=8). This study indicates that there is no significant difference in absorption between two dietary vehicles in rats. That is, absorption of B(a)P from peanut oil and meat was essentially the same. The results with rats and humans also indicates that there is no major difference in the gastrointestinal absorption of B(a)P between rats and humans. Both of the above values were used in AAF derivation.

*Mirvish et al. (1981)*

Mirvish and co-workers (Mirvish *et al.*, 1981) fed B(a)P to Syrian golden hamsters in their diets and measured the amount of unmetabolized B(a)P in their feces to determine the efficiency of absorption from the gastrointestinal tract. B(a)P was dissolved in corn oil, and the corn oil was added to a commercial rodent chow by two different methods. Animals were treated with B(a)P in the diet for 7 to 10 days before samples were collected to give adequate time to reach steady-state PAH concentrations in the feces and gastrointestinal tract contents.

The percentage of fecal excretion of unchanged B(a)P remained relatively constant (94.3% to 98.0%) as its concentration in commercial diet was varied over a wide range (0.16 mg/kg to 5.5 mg/kg). Absorption efficiency was not dose-dependent. The minimal gastrointestinal absorption of B(a)P was found to be 96.7% for the commercial chow using preparation method I (average of results from seven experiments at different dose levels; eleven animal groups, each containing 3-5 hamsters) or 98% for the commercial chow using preparation method II (one experiment; four animal groups, each containing 3-5 hamsters, 1.6 mg/kg). These two values (96.7% and 98%) were used in AAF derivation.

3-methyl cholanthrene (3-MC) absorption was also studied in hamsters. 3-MC (1.7 mg/kg) was dissolved in corn oil and added to a semisynthetic diet consisting of corn oil, corn starch, vitamin-free casein, and alphacel. Minimum gastrointestinal absorption was found to be 93.8% in four animal groups containing 3-5 hamsters each. This value is also used in AAF derivation.

Other experiments demonstrated that B(a)P was absorbed slightly more efficiently from semisynthetic diets than from commercial rodent diets. Addition of corn oil to the hamsters' semisynthetic diets had little effect on the fecal excretion of unchanged B(a)P, and thus its gastrointestinal absorption. Addition of bran to the semisynthetic diets caused a slight lowering of gastrointestinal absorption.

*Rabache et al. (1985)*

Rabache and co-workers (Rabache *et al.*, 1985) fed B(a)P to male Wistar rats in their diets for 22 days and measured the amount of unmetabolized B(a)P in their feces to determine the efficiency of absorption from the gastrointestinal tract. B(a)P was dissolved in soy oil and mixed with the synthetic ration, which was comprised of 10% soy oil. Young rats were given 1 g B(a)P/kg body weight, and adult rats were given 5 g/kg. The minimal gastrointestinal absorption of B(a)P was found to be 88.7% for young rats (n=8) and 99.6% for adult rats (n=12). Both of these values are used in AAF derivation.

*Withey et al. (1991)*

Withey and co-workers (Withey *et al.*, 1991) administered pyrene by stomach tube to male Wistar rats in an aqueous emulsion and measured the amount of C-14 radiolabel in the blood over time to make an estimate of the traditional pharmacokinetic parameter "bioavailability". A single dose of pyrene was given to 4 groups of six animals at a concentration ranging from 4-15 mg/kg as a solution in 20% Emulphor/80% physiological saline. Radiolabeled pyrene was also given intravenously for comparison. "Bioavailability" was defined as the area of the blood level-time curve of radiolabel over a specified time period after oral dosing (0-8 hours) divided by the corresponding area of the curve for intravenous dosing.

"Bioavailability" was found to vary from 65% to 84% depending on dose level. This pharmacokinetic parameter has its basis in classical drug studies where the circulating blood level of the parent (unmetabolized) drug is of primary interest. However, this parameter does not provide an optimal estimate of a chemical's gastrointestinal absorption, because the fraction of the chemical or its metabolites that is bound to tissues is not properly counted.

For this reason, the urinary excretion data over 6 days were also used to derive an estimate of absorption for each group. Absorption was estimated as the fraction of total radiolabel excreted in the urine after oral dosing divided by the fraction excreted after intravenous dosing. Because the fraction excreted in the urine at day 6 post-dosing was slightly higher at every dose level for oral dosing compared to intravenous dosing, the estimates of gastrointestinal absorption are 100% for all four dose groups.

For each dose group, the blood level estimate of "bioavailability" was averaged with the urinary estimate of gastrointestinal absorption to derive an estimate of gastrointestinal absorption. These estimates are: 92%, 82.5%, 86.5%, and 87% for doses ranging from 4-15 mg/kg. The average of these four estimates (87%) is used in AAF derivation.

*Grimmer et al. (1988)*

Grimmer and co-workers (Grimmer *et al.*, 1988) administered chrysene by stomach tube to unfasted male Wistar rats in a solution of 33% dimethylsulfoxide and 66% corn oil. Eight rats weighing 200-250 grams received a single dose of 50 ug chrysene. Assuming an average weight of 225 g, the dose was 0.22 mg/kg. Feces and urine were collected for four days. Unchanged chrysene and specific metabolites were analyzed. The fraction of the unchanged chrysene in the feces was determined. This serves as an estimate of minimal gastrointestinal absorption. Average absorption for the eight rats was 86.9%. This value was used in AAF derivation.

*Bartosek et al. (1984)*

Bartosek and co-workers (Bartosek *et al.*, 1984) administered benz(a)anthracene, chrysene, or triphenylene to female CD-COBS rats by stomach tube in an aqueous emulsion of 10% Pluronic F68 emulsifier and 90% olive oil. Animals were fasted for 24 hours prior to being given a single oral dose of the PAH. Each group consisted of 3-5 rats weighing 150-170 g. PAH were given at single doses of 11.4 and 22.8 mg/ animal, which corresponds to 71.3 mg/kg and 142.5 mg/kg, assuming an average weight of 160 g. Rats were allowed access for food 3 hours after dosing. The fraction of administered dose of the unchanged PAH recovered in the feces after 72 hours was taken as an estimate of the minimal absorption. Results were 94% for benz(a)anthracene, 75% for chrysene, and 97% for triphenylene. These three values were used in AAF derivation.

**Summary of Absorption Data for Dose-Response Studies**

The 13 data points shown in Table 1 are averaged to derive a point estimate of the gastrointestinal absorption of B(a)P and other PAH in the dose-response studies from which the cancer slope factor for B(a)P and the RfDs for various noncarcinogenic PAH were derived. This value is 92%.

Table 1 demonstrates that gastrointestinal absorption of PAHs given in oil vehicles or in the diet is generally high. While there is some variability in the data, no consistent trend is apparent that would lead one to conclude that absorption of one PAH differs significantly from another. Accordingly, all of the data is merged here to represent the absorption of all PAHs of interest.

However, each data point in a study was not given equal weight in deriving the final estimate of oral absorption in the dose-response studies. For instance, in the Mirvish *et al.* study the 96.7% value represents the average of results from seven experiments at different dose levels. There were eleven animal groups, each containing 3-5 hamsters. Thus, this value represents experiments with 33-55 animals. The 98% value represents one experiment at one dose group. There were four animal groups, each containing 3-5 hamsters. Thus, this data point represents 12-20 animals.

There are many ways to summarize such a large and diverse set of experimental results. Table 2, however, demonstrates that the resulting estimate of absorption in the PAH dose-response studies is not particularly sensitive to the manner of summarizing the available data.

## DERIVATION OF ORAL-SOIL AAF FOR POLYCYCLIC AROMATIC HYDROCARBONS (PAH)

Three studies were identified in which the gastrointestinal absorption of PAHs was measured from a soil matrix. These include Goon *et al.* (1991), Rozett *et al.* (1996), and Weyand *et al.* (1996). Each of these studies is discussed below.

*Rozett et al. (1996)*

Rozett *et al.* (1996) studied the bioavailability of pyrene from manufactured gas plant (MGP) residue (coal tar) by comparing the urinary pyrene metabolite levels in animals receiving pyrene as pure MGP residue in their diet to animals receiving pyrene as MGP contaminated soil in their diet. The contaminated soil was aged composite soil from MGP sites. It was fractionated into seven particle size ranges from 1 mm to < 0.150 mm. Soil was added to powder diets from PMI Feeds, Inc. (rodent laboratory diet #5001) (20% soil / 80% powder diet). Pure MGP residue was added to gel diets from Bio-Serv (rodent basal gel diet) (0.003%, 0.03%, 0.1%, & 0.3% coal tar). Groups of female CD1 mice were fed soil or pure MGP residue for 15 days. Urine was collected on day 15. The level of pyrene metabolites (1-hydroxypyrene, 1-hydroxypyrene glucuronide conjugates, and 1-hydroxypyrene sulfate conjugates) were determined by HPLC using fluorescence detection (Singh *et al.*, 1995).

"Bioavailability" is defined by the authors as the amount of pyrene and metabolites excreted in the urine over 24 hours on day 15 divided by the amount of pyrene ingested on day 15 x 100. The amount of pyrene and metabolites excreted into the urine as a fraction of the amount ingested in the last 24 hours is not, itself, a direct measure of bioavailability. It is also not a quantitative measure of total absorption of pyrene from the diet, because PAH and PAH metabolites are efficiently excreted into the feces via the biliary system. However, the level of pyrene and its metabolites in urine on day 15 gives a measure of the steady state level of pyrene excretion. Any pyrene or pyrene metabolite found in the urine necessarily derived from pyrene that was absorbed in the gastrointestinal tract. Because the term bioavailability has a very specific meaning in the fields of toxicology and risk assessment, the metric used by the authors is here renamed "fractional urinary excretion." However, the ratio of "fractional urinary excretion" between study groups is a good measure of relative bioavailability, as will be shown below.

As shown in Table 3, "fractional urinary excretion" of pyrene from MGP residue (coal tar) added to the diet varied from 12.8% to 24.1% depending on the dose level. As shown in Table 4, "fractional urinary excretion" of pyrene from MGP residue-containing soil varied from 1.7% to 14.8% depending on the size fraction of the soil sample. In addition, "fractional urinary excretion" of pyrene from unfractionated soil (< 1 mm particle size) was reported to be 6%.

The ratio of "fractional urinary excretion" from MGP contaminated soil to "fractional urinary excretion" from pure MGP residue as a dietary additive is a direct estimate of the oral-soil AAF (which is a measure of relative bioavailabilty between pyrene in soil and pyrene in food). It is a measure of the degree to which the soil matrix increases or decreases the absorption of pyrene compared to pyrene in the diet. The AAF estimates presented in Table 6 were derived by taking the ratios of "fractional urinary excretion" in Table 4 to the appropriate value from Table 3, based on the dose of pyrene.

*Weyand et al. (1996)*

Weyand *et al.* (1996) studied the bioavailability of pyrene from manufactured gas plant (MGP) residue (coal tar) by comparing the urinary pyrene metabolite levels in animals receiving pyrene

as methylene chloride extracts of MGP contaminated soil in their diet to animals receiving pyrene as MGP contaminated soil in their diet. The two contaminated soil samples were aged soils from MGP sites. They were sieved to a particle size range of less than or equal to 0.150 mm. Soil was added to powder diets from PMI Feeds, Inc. (rodent laboratory diet #5001) (20% soil / 80% powder diet). MGP contaminated soil extracts were added to gel diets from Bio-Serv (rodent basal gel diet) so that the same amount of pyrene was present as in the soil/diet groups. Groups of female $B_6C_3F_1$ mice were fed soil or organic extract for 14 days. Urine was collected on day 14. The level of pyrene metabolites (1-hydroxypyrene, 1-hydroxypyrene glucuronide conjugates, and 1-hydroxypyrene sulfate conjugates) were determined by HPLC using fluorescence detection (Singh *et al.*, 1995).

As above, "fractional urinary excretion" is defined as the amount of pyrene excreted in the urine over 24 hours on day 15 divided by the amount of pyrene ingested on day 15 x 100. The amount of pyrene excreted into the urine is not, itself, a direct measure of total absorption of pyrene from the diet, because PAH are efficiently excreted into the feces via the biliary system. However, the level of pyrene and its metabolites in urine on day 15 gives a measure of the steady state level of pyrene excretion.

As shown in Table 5, the "fractional urinary excretion" of pyrene from soil #1 was 6.2% and from soil #2 was 1.7%. The "fractional urinary excretion" of pyrene from the organic extract of soil #1 was 17.2% and from soil #2 was 16.1%.

The ratio of "fractional urinary excretion" from MGP contaminated soil to "fractional urinary excretion" from an extract of MGP contaminated soil added to diet is a direct estimate of the oral-soil AAF. It is a measure of the degree to which the presence of soil increases or decreases the absorption of pyrene from the diet. The AAF from soil #1 was 36% (6.2%/17.2% x 100). The AAF from soil #2 was 11% (1.7%/16.1% x 100). This study clearly shows that pyrene in aged soil is absorbed in the gastrointestinal tract to a lesser degree than is pyrene added to rodent food as an organic extract.

*Goon et al. (1991)*

Goon, *et al.* (1991) studied the bioavailability of benzo(a)pyrene administered orally as the pure chemical or as B(a)P adsorbed onto soil particles. Additional information about the study was obtained directly from the authors (Goon et al., 1996). Male Sprague-Dawley rats were gavaged with B(a)P mixed with $^{14}$C-B(a)P in solution [0.5% Tween 80 (v/v in saline)] (1.0 μmol B(a)P/kg, 25 μCi/kg) or the equivalent dose adsorbed onto a clay-based soil or a sand-based soil. The soils consisted of 2.5 g solid/kg containing 100 mg/kg B(a)P. All animals received 7.5 mL of 0.5% Tween 80 (v/v in saline).

Venous blood samples were collected from the retro-orbital plexus at predetermined times (0.5, 1, 2, 4, 8, 12, 24, 48, 72, 96, 120, 144, and 168 hours), and excreta were collected continuously over 24-hour intervals. After 168 hours, animals were euthanized and tissues collected for analysis. Total radioactivity was measured by liquid scintillation in blood, urine, feces, and tissues.

The sandy soil was classified as a loam which was very low in organic content, 0.04%. It contained 47% sand, 41% silt, and 12% clay. The pH was 6.5, and the cation exchange content was 0.6 meq/100 g. The clay-based soil was classified as a clay with low organic content, 1.35%. It contained 6% sand, 18% silt, and 76% clay. The pH was 7.0 and the cation exchange content was 45.65 meq/100 g. The sandy soil was ground and sonic sifted. The clay-based soil was dried and passed through a Brickman ultra-centrifugal mill. In both cases, the particles size was small, <100 um. Both soils were washed twice with methylene chloride and dried before use. This destroyed any microbial activity that may have existed in the soils.

B(a)P and $^{14}$C-B(a)P were added in acetone to soils. The acetone was evaporated, leaving soils that were 100 ppm in B(a)P and 10 uCi/g in radiolabel. Animals were administered the soil-adsorbed B(a)P at various time intervals after the soil and the B(a)P were mixed: 1 day, 7 days, 30 days, 6 months and one year. Animals were fasted for 12 hours prior to dosing. Two hours after dosing, Purina Rodent Chow 5001 and water were available *ad libitum*.

Relative bioavailability was measured by comparing the area under the blood curve (AUC) for total radiolabel over the entire 168 hour experimental period during which blood B(a)P levels

were measured. Radiolabel in the blood represents a fraction the B(a)P that was absorbed in the gastrointestinal tract, including parent B(a)P and metabolites.

The use of AUC measurements is a classic approach in drug pharmacology where systemic bioavailability is defined as the blood AUC after an intravenous dose divided by the AUC after an oral dose. In the case of drugs, the amount of parent drug circulating in the blood over a long period of time is of primary interest, because, in most cases, first pass metabolism of the drug in the liver reduces the drug efficacy. Metabolites are inactive and are excreted. Thus, total blood levels of parent drug is of greater interest than is drug plus metabolites.

This same concern is not relevant for the risk assessment of PAHs, such as B(a)P, because B(a)P is not direct acting. No toxic effects are manifested by the parent, unmetabolized B(a)P. Instead, metabolism is required for toxicity. It is the metabolites of B(a)P and other PAH that bind to cellular macromolecules, such as DNA, and cause adverse effects in various tissues. Metabolism of PAHs occurs in all tissues, and orally administered B(a)P has caused tumors in laboratory animals in various tissues, including stomach, lung, esophagus, larynx, and others. B(a)P metabolism is also multistepped. In order for the B(a)P diol epoxide, the putative mutagenic metabolite, to be formed, several metabolic conversions involving several enzymes must occur.

Thus, in some cases the toxic metabolite in a distant tissue, such as the lung, is caused by a B(a)P molecule that was absorbed through the gastrointestinal tract, was *not* metabolized in the liver, circulated through the blood, and was metabolized in several steps in the lung. In other cases, the toxic lung metabolite was formed by a molecule that was absorbed though the gastrointestinal tract, was metabolized to an intermediate metabolite in the liver, and circulated through the blood as a B(a)P metabolite, and was metabolized several more times in the lung to a toxic metabolite.

In addition, B(a)P and B(a)P metabolites excreted in the bile are known to be reabsorbed in the gastrointestinal tract by a process known as enterohepatic recirculation (Chipman et al., 1981). Thus, some B(a)P metabolites are known to be excreted into the bile and the gastrointestinal tract. When present in the gastrointestinal tract parent B(a)P can be reabsorbed. In addition, conjugated metabolites, such as glucuronide, sulfate, and glutathione metabolites can be de-conjugated by enzymes residing in bacteria present naturally in the gastrointestinal tract. After

de-conjugation, the primary metabolite can and is reabsorbed. After reabsorption, it can travel to a distant tissue via the systemic circulation and cause damage.

Thus, for B(a)P and other PAHs, the circulating blood level of just the parent compound is not a relevant dose metric. Instead, the total B(a)P dose including parent B(a)P and metabolites is the critical parameter to measure. This is because some metabolites are directly toxic to distant tissues, some metabolites are metabolic precursors of secondary metabolites that are toxic to distant tissues and can be formed therein, and some metabolites can be excreted and reabsorbed and can later cause damage in distant tissues, including the gastrointestinal tract itself.

While the total blood radiolabel AUC from 0-168 hours does not define the fraction of the administered B(a)P that was absorbed in an animal or a treatment group, the ratio of AUC measurements for two treatment groups administered the B(a)P by the same route of exposure in an excellent measure of *relative* bioavailability between the two treatment groups.

For the clay-based soil, relative bioavailability was 49-59% for the soils that were aged from 1-30 days. For clay-based soils aged 6 months and one year, the relative bioavailability was 39% (see Table 6). For the sand-based soil, relative bioavailability was 67-70% for the soils that were aged from 1-30 days. For clay-based soils aged 6 months and one year, the relative bioavailability was 54% and 62%, respectively (see Table 6).

The above data show that reduction in PAH bioavailability due to soil adsorption is a time dependent phenomenon. This result is consistent with other studies on chemical adsorption to soil. Because the PAH compounds of interest in most soil risk assessments were released to the soil environment many years ago, the results for the 6 month and one year aged soils are used for AAF derivation. These results are 38.6% for clay-based soil and 58.3% for sand-based soil.

These values represent "relative bioavailability" compared to the control animals in which the B(a)P was administered as a solution. They are not direct estimates of gastrointestinal absorption in the soil-treated animals and they are not direct estimates of AAFs. Accordingly, the values must be modified before they can be used to derive AAFs. As shown below, the relative bioavailability value must be multiplied by the absorption in the control animals:

Absorption from soil = Relative Bioavailability x Absorption from solution

The Goon *et al.* (1991) study did not measure total B(a)P absorption in the control animals which received B(a)P in solution. However, four of the absorption estimates presented in Table 1 were from experiments in which the PAH was administered in solution. The results of the five values were averaged to yield 88.5%. Thus, the absorption from sandy soil is estimated as 52% (58.3% x 88.7%). The absorption from clay-based soil is estimated as 34% (38.6% x 88.5%). The AAFs are defined as the absorption from soil divided by the absorption from diet x 100. They are as follows:

AAF oral-soil (sandy) = 52% / 92% = 0.57
AAF oral-soil (clay-based) = 34% / 92% = 0.37

Ogden notes that the two soils studied were very low in organic content (0.04% and 1.35%). Certainly, the value for sandy soil is much lower than a typical soils. For instance, in its Risk Based Corrective Action guidance, the ATSM assumes 1% as a default value for typical soils. Accordingly, the AAF for clay-based soil is probably more typical of average soils than the AAF for sandy soil.

*Goon et al. (1990)*

In an earlier experiment, Goon et al. (1990) studied the bioavailability of B(a)P in aqueous solution, in laboratory chow, in unaged sandy soil and in unaged clay-based soil. Additional information was obtained directly from the authors (Goon et al., 1996). The study was performed in the same manner as the one described above with the exception that 4 male rats and 4 female rats were placed in each of four study groups, including rodent chow.

In that study, the bioavailability from rodent food was shown to be less than from solution. When the area under the curve for total radioactivity in blood over 168 hours was compared, the solution group was 5944 pmol-hour/mol and the rodent chow group was 3179 pmol-hour/mol. Thus, bioavailability from food was 54% compared to aqueous solution. Bioavailability of B(a)P administered in slurries adsorbed onto small particles from sand and clay-based soils were also decreased relative to B(a)P in solution (47% for sandy soil and 28% for clay-based soil).

Ogden has rejected the data from the Goon et al. (1990) study for AAF derivation and relied solely on the 1991 experiment for several reasons. First, the results for B(a)P adsorbed to rodent chow and dissolved in a solution with an aqueous emulsifier are at variance from the results presented in the large literature on B(a)P absorption discussed above. Table 1 shows that in all other studies of B(a)P and other PAHs, absorption is high and similar for PAHs adsorbed to food (either meat or rodent chow), dissolved in vegetable oils, or dissolved in emulsifier solutions.

Second, the results for each treatment group were averaged over data for both males and females, which had very different starting and ending body weights (see Table 7). The starting body weight for female rats was 75% to 81% of the body weight of the male rats. Goon et al. in the 1990 experiment averaged the blood radioactivity levels for 3-5 male and 4-5 female rats in each treatment group and then calculated a group-wide area under the curve (AUC). They did not calculate the AUC for the total 168 hour experiment for each animal and then average the animal-specific AUC's. Thus, a sex-specific reduction in bioavailability or any source of animal-specific variability could lead to artifacts in the group average AUCs.

Third, Ogden has uncovered such variability by evaluating the data for body weights and the weight gain over the experimental period. Table 7 shows the weights of the animals in each group before the 12 hour fast period, after the fast period and before dosing, and after the 7 day experiment. Ogden notes that the variability in the weights of the male animals in the solution group and in the sandy soil group is much higher than the variability in any of the other groups. In particular, the variabilities in the post experiment weights for animals in the food groups (male and female) were much smaller than the variability in the male solution group. (The variability in animal weights in all groups, including the solution group, was much smaller in the 1991 experiment (see Table 9)).

Lastly, when the pre-experiment and post-experiment animal weights (see Table 7) are compared, an interesting result is apparent. The weight gain over the experimental period is much higher for the solution group than the food group (see Table 8). On average, the males in the solution group gained 9% of their initial fasted weight. The females gained 8% of their initial fasted weight. However, the males and females in the food group gained 15% and 17% of their initial fasted weight, respectively. In the Goon et al. (1991) experiment, the B(a)P-solution group, on

average, gained 15% of their initial fasted weight (see Table 9), again showing that the results from the 1990 experiment are suspect.

A reasonable explanation for the anomalous results for B(a)P absorption from food and solution doses is that, for some unknown reason, the animals in the solution group consumed less food immediately after the B(a)P dosing than did the group that received B(a)P in a slurry of rodent chow. If the animals in the food group ate more food, then the B(a)P was diluted by a large volume of soil and water with a greater surface area of material to which it could bind, preventing gastrointestinal absorption. With the solution group, if they ate less food following the dosing, then the B(a)P present in their empty stomachs in an emulsified aqueous solution could be rapidly absorbed, perhaps quantitatively.

Ideally, one could test this hypothesis by studying food consumption records. However, food was provided *ad libitum*, and daily animal-specific food consumption was not monitored. However, the weight gain over the period is a rough measure of food consumption.

The 1990 experiment is also suspect when one compares the male weight gains and the female weight gains in each treatment group. In the solution, food, and clay-based soil groups, the males gained more weight over the experimental period (10-18% more than the females). However, in the sandy soil group, the females gained 91% more weight than the males. Clearly, the results for sandy soil are suspect.

Ogden does not know why certain groups would have consumed more food and gained considerably more weight than others. Perhaps radiation-induced or emulsifer-induced gastritis or diaharrea was the cause. Although all groups received 7.5 mL of emulsifer, in the solution group, this was the only material administered on an empty stomach. In the other four groups, this was given with 2.5 grams of solid material, which would have been wetted by the solution.

Regardless of the reasons for the inadequacies of the 1990 study, the 1991 experiment does not suffer from these sorts of variabilities and differences in weight gains. In addition, the experiment used only male animals, so the uncertainties and confounding effects of averaging the results over animal groups with widely differing body weights and food consumption rates

are not seen. Accordingly, Ogden used the data from Goon et al. (1991) in AAF derivation, but rejected the use of the Goon et al. (1990) data.

**Summary of Oral-Soil AAFs**

Twelve estimates of the oral-soil AAF for PAHs were derived from three studies, as shown in Table 10. For probabilistic risk assessments, a distribution of AAF values is required. Curve fitting exercises using Mathematica™ software and using the methods shown in Burmaster (1996) determined that the 12 data points best fit a Beta4 distribution with the following characteristics: Beta4 (a=1, b=3, c=0.944964, d=0.0699) over the range of 0.07-1.00. Then, Monte Carlo simulations were run using Crystal Ball™ software. The mean oral-soil AAF for PAHs after 20,000 trials was 0.31 with a standard deviation of 0.18. The 50th percentile oral-soil AAF was 0.27 and the upper 90th percentile oral-soil AAF was 0.57. For deterministic risk assessments, a point estimate is needed for the AAF. The average of the twelve values is 0.29. This average value is similar to the mean and 50th percentile values from the AAF distribution. Accordingly, 0.29 is an appropriate point estimate of the oral-soil AAF.

**Applicability of Oral-Soil AAFs**

These estimates of oral-soil AAFs were derived from studies with B(a)P, a five-ring potentially carcinogenic PAH and pyrene, a four-ring noncarcinogenic PAH. Because the AAF estimates for the two PAHs were similar and because the gastrointestinal absorption of various potentially carcinogenic and noncarcinogenic PAHs is similar (see Table 1), it is appropriate to derive a single oral-soil AAF for the carcinogenic and noncarcinogenic risk assessment of all potentially carcinogenic PAHs.

## DERIVATION OF DERMAL-SOIL AAF FOR POTENTIALLY CARCINOGENIC POLYCYCLIC AROMATIC HYDROCARBONS (PAH)

Two studies were identified in which the dermal absorption of PAHs was measured from a soil matrix. These include Yang *et al.* (1989) and Wester *et al.* (1990). These studies are discussed below. Estimates of dermal-soil AAFs can be derived from the results of these studies and data on absorption from the dose-response studies.

**Dermal Absorption Studies**

*Yang et al. (1989)*

Yang, *et al.* (1989) measured the percutaneous absorption of benzo(a)pyrene (B(a)P) from petroleum crude-fortified soil and from pure petroleum crude oil both in live rats and in *in vitro* studies using excised rat skin (see Table 11). The soil was a loam containing 1.64% organic matter, 46% sand, 36% silt, and 18% clay. The B(a)P-soil mixture was prepared by adding the radiolabelled crude oil in dichloromethane to the soil. The solvent was removed by rotary evaporator. All soils were used within 72 hours of preparation.

Radiolabelled B(a)P ($^3$H-B(a)P) was added at a known concentration for quantification. In the *in vivo* experiments, soil containing B(a)P in crude petroleum or pure crude petroleum containing B(a)P was applied to the dorsal skin of the female Sprague-Dawley rats. In both cases, the dose of B(a)P was 0.01 ug/cm$^2$. For the crude oil, 90 ug/cm$^2$ of oil containing 100 ppm B(a)P was applied. For soil, 9 mg/cm$^2$ of soil containing 1 ppm of B(a)P was applied. The dorsal area was covered with a non-occlusive glass cell to prevent ingestion of the B(a)P by grooming behavior.

Absorption was determined by measuring the radioactivity in the urine and feces once daily and the urine, feces and tissues at 96 hours. Data from five animals were averaged. After 96 hours, cumulative absorption of B(a)P from crude-soaked soil (9.2%) was less than that from the crude alone (35.3%).

In the *in vitro* experiments, dorsal skin was excised from female Sprague-Dawley rats after sacrifice. 350 um skin sections were placed in consoles containing 15 mm diameter Franz diffusion cells. The receptor fluid was an aqueous solution of 6% Volpo-20, a nonionic surfactant. The absorption was measured by analyzing the surfactant containing receptor fluid that bathed the receiving reservoir of the absorption chamber for radiolabelled B(a)P. The receptor fluid was sampled once every 24 hours for four days. Data from five trials were averaged. Again, 96 hour cumulative absorption was greater for B(a)P in oil (38.1%) versus B(a)P in oil-soaked soil (8.4%).

*Wester et al. (1990)*

Wester *et al.* (1990) measured the absorption of B(a)P *in vivo* over 24 hours in the monkey using acetone as vehicle or using soil containing B(a)P at the 10 ppm level (see Table 12). The soil used contained 26% sand, 26% clay, and 48% silt. The organic content was not specified. The B(a)P containing soil was prepared by adding the B(a)P in (7:3, v/v) hexane:methylene chloride. The soil was mixed by hand and left open to the air to allow dissipation of the solvent. The B(a)P-soil mixture was not aged before use.

Four female Rhesus monkeys were tested with 40 mg soil/cm$^2$ applied to the abdominal skin. The skin area was covered with a nonocculusive cover to prevent loss of soil or ingestion of soil by grooming behavior. Percutaneous absorption was measured by comparing the quantity of radiolabel ($^{14}$C-B(a)P) in the urine following topical application to that following intravenous application. Urine was collected for 24 hours. After 24 hours, all visible soil was collected from the application site. The skin surface was washed with soap and water, and the monkeys were returned to metabolic cages for urine collection for an additional six days. *In vivo*, the absorption was 51.0% for acetone vehicle and 13.2% for soil.

*In vitro* studies were also carried out with viable human cadaver skin in cells of the flow-through design. Human serum was used as the receptor fluid. Radiolabel was determined in the receptor fluid after 24 hours as well as in the skin after a surface wash with soap and water. The amount of B(a)P that cannot be removed from the skin with a soap and water wash is designated here as "absorbed" for the purposes of AAF derivation. In six experiments with skin from two donors, 23.8% of the B(a)P was absorbed with acetone vehicle. From soil (10 ppm), 1.45% was absorbed in 24 hours.

**Dermal-Soil AAF Derivation**

The fraction absorbed in a 24-hour or 96-hour experiment has little relevance to human risk assessment. People who might touch, walk on, or otherwise contact PAH-containing soil would only be exposed for a period of 6-12 hours at maximum before washing themselves or before the soil would drop off or be rubbed off the skin. The Wester, *et al.* (1990) paper demonstrates that

soap and water wash can remove a large amount of the administered dose (53-91%), even after 24 hours. Even more would be removed after only 6-12 hours exposure.

EPA guidance for dermal risk assessment recognizes that the time period of a dermal experiment is an important factor to consider when evaluating experimental data. EPA (1992b) has noted: "The experiment should provide absorption estimates over a time corresponding to the time that soil is likely to remain on skin during actual human exposures."

Accordingly, the data from the Yang, *et al.* (1989) and Wester, *et al.* (1990) experiments should be prorated for a reasonable exposure period, such as 6-12 hours. A health-protective way to do this is to simply assume that absorption is linear over time. The Yang, *et al.* (1989) *in vitro* study showed a linear absorption into rat skin from 24-96 hours, but no data are available for the 0-24 hour period.

In fact, Kao *et al.* (1985) have shown that the appearance of radiolabel from topically applied benzo(a)pyrene and other chemicals in human, rodent, and other species' skin in the culture medium of their *in vitro* system was exponential, not linear. A distinct time lag is apparent before any absorption occurs. A time lag has also been shown for various chlorophenols in human skin (Roberts, *et al.*, 1977; Huq, *et al.*, 1986). EPA (1992b) also recognizes that a time lag may exist: "time is required after initial contact with the skin for such a steady-state to be achieved." Also: "Linear adjustments may not be accurate, since it is unknown how soon steady-state is established and since steady-state conditions may not be maintained throughout the experiment due to mass balance constraints."

Thus, linear adjustments of 24 hour absorption data to estimate absorption over 6-12 hours may overestimate the absorption true absorption, but it is not likely to underestimate absorption. A health-protective approach would be to assume that a relevant absorption period is as high as 12 hours. (EPA in its recently proposed Hazardous Waste Identification Rule assumes 8 hour exposures.) With this assumption, the Yang *et al*, 1989 data from the *in vitro* experiment can be adjusted to 0.66% absorption over 12 hours using a linear regression of all four time points. The data from the *in vivo* experiment can be adjusted to 1.15% absorption over 12 hours. The 96 hour data is used in this case, because tissue-bound B(a)P was measured only for this time

point. The 12 hour estimated absorption using a linear regression is only 0.50%, and was thus rejected for AAF derivation. .

The Wester, *et al.* (1990) data can be adjusted to 6.6% absorption in the *in vivo* monkey experiment over a 12 hour exposure period. Similarly, the 12 hour estimated exposure for the *in vitro* human skin experiment is 0.73%.

For probabilistic risk assessments, a distribution of AAF values is required. The numerator and the denominator of the AAF ratio are defined as separate distributions which are sampled independently during the probabilistic risk assessment.

Curve fitting exercises for the numerator (dermal absorption of potentially carcinogenic PAHs from soil) using Mathematica™ software and the methods described in Burmaster (1996) indicated that the four data points best fit a Beta4 distribution with the following characteristics: Beta4 (a=1, b=5, c=0.146908, d=0) over the range 0-0.12. Monte Carlo simulations were then run using Crystal Ball™ software. The mean fractional dermal absorption of potentially carcinogenic PAHs after 20,000 trials was 0.02 with a standard deviation of 0.02.

Curve fitting exercises for the denominator (gastrointestinal absorption of PAHs from dose-response studies) using Mathematica™ indicated that the 13 data points for absorption in the PAH dose-response studies best fit a Beta distribution with the following characteristics: Beta4 (a=4, b=1, c=0.397, d=0.602697) over the range 0.63-1.00. Monte Carlo simulations were then run using Crystal Ball™. The mean fractional gastrointestinal absorption of PAHs in the dose-response studies after 20,000 trials was 0.92 with a standard deviation of 0.06.

Monte Carlo simulations of the dermal-soil AAF were then run using these assumptions. The mean dermal-soil AAF for potentially carcinogenic PAHs after 20,000 trials is 0.03 with a standard deviation of 0.02. The 50th percentile AAF was 0.02, and the 90th percentile AAF is 0.06.

For deterministic risk assessments, a single estimate of the dermal-soil AAF is needed. In this case, four estimates of the dermal absorption of PAHs from soil were presented: 0.66%, 0.73%, 1.15%, and 6.6%. In addition, 12 estimates of the absorption of PAHs from the dose-response

study were presented in Table 1. The average value is 92%. Four AAF estimates are 0.007, 0.008, 0.01, and 0.07. The deterministic estimate of the dermal-soil AAF is simply the average of the four AAFs, 0.02. This value is similar to the mean and 50th percentile estimates for the AAF distribution, and is thus appropriate for use in deterministic risk assessments.

## APPLICABILITY OF DERMAL-SOIL AAF TO OTHER CARCINOGENIC PAHS

Dermal-soil AAFs have been derived for B(a)P based on four experimental data points with B(a)P. However, risk assessment of PAHs involves the calculation of benzo(a)pyrene-toxic equivalents, which includes the seven PAHs designated as potentially carcinogenic. The following section addresses the applicability of the B(a)P AAF to other potentially carcinogenic PAHs.

Various researchers have investigated the dermal absorption of different PAHs from pure mixtures, such as coal tar, or from solvent vehicles, such as acetone. From these studies, data on the comparative dermal absorption of various pure PAHs are available, but no studies are available on the dermal absorption of various PAHs from a soil matrix.

For instance, Sanders, *et al.* (1984) studied the dermal absorption of B(a)P and dimethylbenz(a)anthracene (DMBA) in Swiss-Webster mice from an acetone vehicle. The dermal absorption was similar for the two PAHs. For instance, at similar dose levels, the amount found in the tissues and excreta 24 hours after dosing was 84% for B(a)P and 82% for DMBA.

Yang and coworkers (Yang *et al.* 1986a, 1986b) studied dermal absorption of B(a)P and anthracene at similar doses from solvent vehicles in the female Sprague-Dawley rat in both *in vivo* and *in vitro* systems. Absorption was similar for the two PAHs. *In vivo*, absorption after 144 hours was 46.2% for B(a)P and 52.3% for anthracene. *In vitro*, absorption after 144 hours was 49.9% for B(a)P and 55.9% for anthracene.

Ng and coworkers (Ng *et al.*, 1992) studied dermal absorption of B(a)P and pyrene at similar doses from an acetone vehicle in the hairless guinea pig. Absorption after 24 hours was 73.3% for B(a)P and 93.9% for pyrene. In an *in vitro* experiment, absorption of B(a)P was 67.4%

versus 89.9% for pyrene. In another *in vitro* experiment, absorption of B(a)P was 39.8% versus 40.8% for pyrene.

Dankovic and colleagues (Dankovic *et al.*, 1989) studied the comparative dermal absorption in female CD-1 mice of 12 high molecular weight PAHs isolated from the 800-850 degree (F) complex organic mixture (COM) derived from a coal liquefaction process. Absorption was measured as the half life of disappearance of the PAH from the mouse skin. The half life was 5.0 hours for pyrene. For B(a)P, the half life was 6.7 hours. All other PAH had half lives similar to B(a)P, including benz(a)anthracene (6.5 hr), chrysene (7.3 hr), and benzo(j/k)fluoranthene (8.1 hr).

VanRooij *et al.* (1995) studied the dermal absorption in the blood-perfused pig ear of 10 PAHs present in coal tar. The blood-perfused pig ear was chosen as a test system because pig skin resembles human skin morphologically and functionally and because percutaneous absorption rates of various chemicals in pig skin are comparable to the rates seen in human skin.

The absorption after 3.3 hours varied among PAHs. Absorption was greatest for phenanthrene and fluorene. Anthracene, fluoranthene, and pyrene showed similar absorption rates that were roughly ten times less than those for phenanthrene and fluorene. The 4-6 ring PAHs showed substantially lower dermal absorption, which was 100-1000 times less than that seen with phenanthrene and fluorene. It should be noted, however, that the maximum fractional absorption seen, which was with fluorene, was only 0.004% of the applied dose.

Of the potentially carcinogenic PAH studied in the above dermal absorption experiments, B(a)P showed equal or greater dermal absorption. None of these experiments were performed with soil matrices. They all involved applying the PAHs as solutions in organic solvents.

As noted above, dimethylbenz(a)anthracene, benz(a)anthracene, and benzo(b)fluoranthene were absorbed to a degree similar to B(a)P. Chrysene, benzo(k)fluoranthene, indeno[1,2,3-cd]pyrene, and dibenzo(a,h)anthracene were absorbed to a lesser degree than was B(a)P. Accordingly, it is health protective to use dermal-soil AAFs derived for B(a)P for performing risk assessment of all potentially carcinogenic PAH.

## RELEVANCE TO HUMAN ABSORPTION

Limited quantitative data are available on PAH absorption in humans. By the oral route, absorption of pure B(a)P was shown in one study to be similar in humans compared to that seen in rats and hamsters. However, no data are available on the human gastrointestinal absorption of PAHs in a soil matrix. The literature presents no basis for presuming that gastrointestinal absorption of PAHs from soils would be significantly different in humans and experimental animals.

By the dermal route, several studies are available that document absorption of PAHs from pure mixtures, such as coal tar, in human subjects. For instance, Clonfero *et al.* (1986) measured PAH metabolites in the urine of humans dermally exposed to coal tar. Storer *et al.* (1984) measured PAH levels in the blood of humans exposed to coal tar. Finally, Schoket *et al.* (1990) measured aromatic DNA adducts in the skin of humans exposed to coal tar. These and other studies clearly demonstrate that absorption of PAHs from pure mixtures or from PAHs dissolved in solvents can occur in human skin.

Only three, however, are available that have quantitated the absorption of pure PAHs or PAHs in soil matrices in human skin. As discussed above, Wester *et al.* (1990) studied the absorption of B(a)P in an acetone vehicle and in soil in both monkeys and in human skin *in vitro*. The absorption from acetone was 2.1 times higher over 24 hours in the monkey compared to the human skin. From the soil matrix, absorption was 9.1 times higher in the monkey compared to the human skin.

Kao *et al.* (1985) studied the absorption of B(a)P from acetone in an *in vitro* system with skin from six species, including humans. Absorption over 24 hours was highest in the mouse. Absorption in the marmoset, rat and rabbit was similar to that in human skin. Absorption in the guinea pig was the lowest.

Storm *et al.* (1990) studied the absorption of B(a)P *in vitro* in flow through diffusion cells with skin from humans, two rat strains, guinea pig, and two mouse strains. Absorption over 24 hours was similar in the mice, rats, and guinea pig. Absorption in human skin, however, was significantly lower by 1.5-2 fold.

Available studies indicate that human skin is less permeable to PAHs in pure form than is rodent or monkey skin. Thus, the dermal-soil AAF may overestimate the true AAF for human skin. Because the dermal-soil AAFs are derived from data on rats, monkeys, and humans, however, they are reasonable, health-protective estimates for use in human health risk assessment.

## SUPPORTING EVIDENCE THAT SOIL ADSORPTION REDUCES GASTROINTESTINAL AND DERMAL ABSORPTION OF PAH

There are several bodies of experimental data that support the concept that soil adsorption over time binds and sequesters PAH molecules so that they are unavailable for absorption in the skin and gastrointestinal tracts of humans and animals that might contact the affected soils. The results of these experiments cannot easily be used to derive a quantitative estimate of the lowering in absorption, but they are presented here as scientific justification of the phenomenon.

### Studies on Soil Bioavailability of other Chemicals

Several studies were identified that compared tetrachlorodibenzodioxin (TCDD) absorption from soil to either diet, oil vehicle, or alcohol vehicle. These studies demonstrate that gastrointestinal absorption of TCDD is reduced when present as a component of soil or other matrix that can adsorb the TCDD. Dioxins and PAHs are two classes of lipophilic chemicals that would be expected to behave similarly with regard to soil adsorption.

For instance, Van den Berg and co-workers (1983) administered PCDDs and PCDFs from fly-ash and fly-ash extract to male Wistar rats as a dietary constituent. The absorption from fly ash was only 22% of the absorption from extracts.

Other studies are available in which absorption of TCDD from soil was compared to oil or alcohol vehicles. McConnell *et al.* (1984) investigated absorption in guinea pigs using soil from Missouri that contained TCDD. Gastrointestinal absorption from soil was 15-24% of the absorption from corn oil.

In a similar experiment, Poiger and Schlatter (1980) studied the effects of soil adsorption on the oral bioavailability of TCDD in Sprague-Dawley rats. When TCDD was administered as an

aqueous suspension of soil particles that had been in contact with the TCDD for 8 days, the fraction of the administered dose that was found in the liver 24 hours later was 43% of that found with an aqueous ethanol vehicle.

Similar studies have also been performed in rabbits by Bonaccorsi *et al.* (1984). Levels of TCDD in the liver 7 days after an oral dose of TCDD either in alcohol or in soil from Seveso, Italy were compared. The ratio of TCDD absorption from soil relative to alcohol vehicle was 32% in this study.

Umbreit *et al.* (1986) also studied the effect of soil adsorption on 2,3,7,8-TCDD-induced toxicity in guinea pigs. Dioxin as a suspension of corn oil and acetone (9:1) (6 ug/kg) given to guinea pigs by stomach tube caused death in 5 of 8 animals within 5-31 days, and autopsy showed signs typical of the TCDD-induced toxicity that is observed in the guinea pig. When the same amount of 2,3,7,8-TCDD was placed on soil for only one hour and then administered to the animals, similar results were seen. However, contaminated soil from a site in New Jersey containing the same or double the amount of 2,3,7,8-TCDD failed to cause any deaths and also failed to induce any recognizable signs of TCDD-induced toxicity. Thus, aging of the soil causes decreased bioavailability.

**Studies on Effects of Dietary Components on PAH Absorption**

Several studies have been evaluated on the effects of dietary fiber and other food items on PAH absorption in the gastrointestinal tract. In general, it has been shown that dietary fiber of various types can bind or adsorb PAH and reduce their absorption in the gut of experimental animals. For instance, Gulliver *et al.* (1983) showed that dietary fiber binds dimethylbenz(a)anthracene *in vitro* and decreases solubilization by bile salt solutions by 61-98%. Mirvish et al. (1981) showed that B(a)P absorption in rats was reduced from 99.8% in semisynthetic diets having no fiber to 95% when wheat bran was added. Kawamura *et al.* (1988) studied B(a)P absorption from various food items in the rat. Absorption was highest when B(a)P was administered in triolein oil. When B(a)P was given in different food items that included cellulose, bread, lignin, ovalbumin, spinach, and others absorption was reduced to as low as 40% of that seen with triolein. Similar results were seen with the release of B(a)P from food items *in vitro* in artificial intestinal fluid.

**Studies on the Effects of Soil Components on PAH Mutagenicity**

Sato *et al.* (1987) studied the effects of organic chemicals found in soil on the mutagenicity of B(a)P to *Salmonella typhimurium.* Humic acid and lignin totally inhibited the ability of B(a)P to mutate the bacteria in culture. Fulvic acid and water-soluble humic substances inhibited B(a)P - induced mutagenicity to a lesser degree. It was found that the humic acid inhibited mutagenicity by binding the B(a)P and making it unavailable to the bacteria in culture. This was shown by mixing B(a)P and humic acid and then extracting the B(a)P by ethyl acetate. In the presence of humic acid only 25% of the B(a)P could be extracted compared to controls containing no humic acid. All of the added B(a)P could, however, be released after ultrasonication, indicating that the humic acid was reducing B(a)P's bioavailability.

**Studies on Solvent Extractability of PAH from Soils**

Karickhoff (1980) showed that PAHs became increasingly more difficult to extract from sediments with increasing contact time. For instance, after 4 minutes pyrene was 94% recoverable with solvent extraction, but after 122 hours only 36% could be recovered. Quantitative recovery after a 72 hour Soxhlet extraction confirmed that the PAH had not degraded, but rather was adsorbed tightly to sediment particles.

Hatzinger and Alexander (1995) showed that butanol extractability of phenanthrene decreased from 95% recovery to 61% recovery from a high organic content soil when the mixture was aged 84 days. The soil was sterilized to prevent bacterial degradation. Greater recoveries after Soxhlet extraction confirmed that soil adsorption was the reason for reduced solvent extraction efficiency.

**Studies on Bacterial Degradation of PAH in Soils**

Hatzinger and Alexander (1995) introduced phenanthrene into high organic content soils that had been sterilized to remove organisms that might degrade the PAH. After aging the phenanthrene in the soil for varying periods of time (29 weeks, 45 weeks), a phenanthrene-degrading organism was introduced. After a month, 60% of the phenanthrene was degraded in the unaged control. Bacterial degradation was diminished in the aged soils. Degradation plateaued at 45% for the

29 week soil and at 40% for the 45 week soil. Adsorption of the PAH to the soil was responsible for the reduction in its bioavailability to microorganisms.

Weissenfels *et al.* (1992) studied the biodegradation of PAHs in soils from a closed coking plant. PAHs were not degraded by autochthonous organisms or after inoculation with bacteria known to degrade PAHs. However, rapid degradation of PAHs was observed when PAHs were extracted from the soil by an organic solvent and then re-introduced into the extracted soil material. Sorption of the extracted PAHs onto the extracted soil followed a two-phase process. The authors described the slow phase of sorption as migration into less accessible sites within the soil matrix. The authors concluded that the PAHs so sorbed within the soil matrix are non-bioavailable and non-biodegradable. The initial soil was extracted with water and assayed for toxicity with bioluminescent bacteria. No toxicity was observed in the aqueous phase.

**Studies on Reduction in Chemical Toxicity after Aging in Soil Matrices**

Edwards *et al.* (1957) showed that the lethal dose of lindane and aldrin in *Drosophila melanogaster* increased as soil organic content increased. The $LD_{50}$ for lindane varied from 0.25 mg/kg in soils containing 0.5% organic matter to 8.6 mg/kg in soils containing 40% organic matter. For aldrin, the results were similar. Peterson *et al.* (1971) reported a similar result for DDT in *Drosophila melanogaster.* The $LD_{50}$ increased from 43 to 790 mg/kg as the fraction of organic matter in the soil increased.

## SUMMARY

The point estimate Oral-Soil AAF derived for deterministic risk assessment of potentially carcinogenic PAH is 0.29. For probabilistic risk assessments, the Oral-Soil AAF distribution is defined as a Beta4 distribution with the following characteristics: Beta4 (a=1, b=3, c=0.944964, d=0.0699) over the range of 0.07-1.00.

The point estimate Dermal-Soil AAF derived for deterministic risk assessment of potentially carcinogenic PAH is 0.02. For probabilistic risk assessments, a distribution of Dermal-Soil AAF values is required. The numerator and the denominator of the AAF ratio are defined as separate distributions which are sampled independently during the probabilistic risk assessment. The numerator (dermal absorption from soil) is defined as a Beta4 distribution with the following characteristics: Beta4 (a=1, b=5, c=0.146908, d=0) over the range 0-0.12. The denominator (gastrointestinal absorption of PAHs from dose-response studies) is defined as a Beta4 distribution with the following characteristics: Beta4 (a=4, b=1, c=0.397, d=0.602697) over the range 0.63-1.00.

## TABLE 1

## SUMMARY OF ABSORPTION DATA FOR PAH DOSE-RESPONSE STUDIES

| Value | Citation | Animal | PAH | Vehicle |
|---|---|---|---|---|
| 91.2% | Hecht | male F344 rats | B(a)P | peanut oil (single dose) |
| 89% | Hecht | male F344 rats | B(a)P | char-broiled hamburger (single dose) |
| 98.8% | Hecht | Humans | B(a)P | char-broiled hamburger (single dose) |
| 88.7% | Rabache | young male Wistar rats | B(a)P | synthetic diet (22 days) |
| 99.6% | Rabache | adult male Wistar rats | B(a)P | synthetic diet (22 days) |
| 96.7% | Mirvish | male Syrian golden hamsters | B(a)P | corn oil + commercial diet Method I (7-10 days) |
| 98.0% | Mirvish | male Syrian golden hamsters | B(a)P | corn oil + commercial diet Method II (7-10 days) |
| 87% | Withey | male Wistar rats | pyrene | 20% Emulphor/ 80% saline (single dose) |
| 86.9% | Grimmer | male Wistar rats | chrysene | 33% DMSO/ 66% corn oil (single dose) |
| 94% | Bartosek | female CD-COBS rats | B(a)A | 10% emulsifier/ 90% olive oil (single dose) |
| 75% | Bartosek | female CD-COBS rats | chrysene | 10% emulsifier/ 90% olive oil (single dose) |
| 97% | Bartosek | female CD-COBS rats | triphenylene | 10% emulsifier/ 90% olive oil (single dose) |
| 93.8% | Mirvish | male Syrian golden hamsters | 3-methyl cholanthrene | corn oil + semisynthetic diet (7-10 days) |

## TABLE 2

## METHODS OF SUMMARIZING PAH GASTROINTESTINAL ABSORPTION DATA

| Method Used | # Data Points | Average Absorption |
|---|---|---|
| Each experiment within a study used as a single data point* | 13 | 92.0% |
| Each result presented in each study used as a single data point | 24 | 92.1% |
| Each result presented in each B(a)P study used as a single data point | 15 | 95.0% |
| Each study represented as a single data point | 7 | 90.9% |
| Each B(a)P study represented as a single data point | 3 | 94.4% |

* Method used in this AAF derivation.

**TABLE 3**
**PYRENE METABOLITES IN MOUSE URINE**
**FOLLOWING "NEAT" MGP INGESTION**
**(ROZETT *ET AL.*, 1996)**

| Amount of MGP residue in diet | [a]Sum of Metabolites µg/mouse | [b]Pyrene consumed µg/mouse | [c]Fractional Urinary Excretion |
|---|---|---|---|
| 0.003% | 0.10 | 0.79 | 12.8 |
| 0.030% | 1.39 | 11.39 | 12.2 |
| 0.100% | 7.58 | 31.46 | 24.1 |
| 0.300% | 12.13 | 62.27 | 19.5 |
| Control | - | - | - |

[a]The sum of 1-OH P-GlcUA, 1-OH P-Sul, and 1-OH P levels is expressed in terms of equivalents of pyrene.
[b]The amount of pyrene consumed by animals in metabolism cages on day 15 over a period of 24 hours.
[c]Fractional Urinary Excretion = (amount of pyrene excreted / amount of pyrene consumed on day 15) x 100. (The authors termed this "bioavailability." Because this is a nonstandard use of the term, it is renamed here.)
Note: The pyrene level in "neat" MGP was 6.89 mg/kg.

## TABLE 4

## PYRENE METABOLITES IN MOUSE URINE FOLLOWING SOIL INGESTION (ROZETT *ET AL.*, 1996)

| Soil Fraction | [a]Sum of Metabolites µg/mouse | [b]Soil consumed g/mouse | Pyrene in soil µg/g | [b]Pyrene consumed µg/mouse | [c]Fractional Urinary Excretion |
|---|---|---|---|---|---|
| >0.850 mm | 0.37 | 0.65 | 14.3 | 9.4 | 3.9 |
| >0.710 mm | 0.69 | 0.64 | 61.8 | 39.7 | 1.7 |
| >0.600 mm | 0.70 | 0.68 | 63.4 | 43.1 | 1.6 |
| >0.500 mm | 0.95 | 0.63 | 74.6 | 47.2 | 2.0 |
| >0.300 mm | 1.72 | 0.66 | 26.8 | 17.7 | 9.7 |
| >0.150 mm | 1.77 | 0.58 | 177.9 | 102.4 | 1.7 |
| ≤0.150 mm | 9.86 | 0.36 | 185.6 | 66.7 | 14.8 |
| Control | - | - | - | - | - |

[a]The sum of 1-OH P-GlcUA, 1-OH P-Sul, and 1-OH P levels is expressed in terms of equivalents of pyrene.
[b]The amount of soil and pyrene consumed in metabolism cages on day 15 over a period of 24 hr.
[c]Fractional Urinary Excretion = (amount of pyrene excreted / amount of pyrene consumed on day 15) x 100. (The authors termed this "bioavailability." Because this is a nonstandard use of the term, it is renamed here.)

## TABLE 5

## PYRENE URINARY METABOLITES

## SOIL VS ORGANIC EXTRACT OF SOIL (WEYAND *ET AL.*, 1996)

| Diet | [a]Pyrene Ingested (µg/mouse) | [b]Pyrene Excreted (µg/mouse) | [c]Fractional Urinary Excretion |
|---|---|---|---|
| Extracted Soil #1 | 0 | 0 | ND |
| Extracted Soil #2 | 0 | 0 | ND |
| Soil #1 | 0.60 | 0.039 | 6.2 |
| Soil #2 | 30.42 | 0.527 | 1.7 |
| Organic Extract #1 | 0.56 | 0.097 | 17.2 |
| Organic Extract #2 | 25.91 | 4.16 | 16.1 |

[a]The sum of 1-OH P-GlcUA, 1-OH P-Sul, and 1-OH P levels is expressed in terms of equivalents of pyrene.
[b]The amount of soil and pyrene consumed in metabolism cages on day 15 over a period of 24 hr.
[c]Fractional Urinary Excretion = (amount of pyrene excreted / amount of pyrene consumed on day 15) x 100. (The authors termed this "bioavailability." Because this is a nonstandard use of the term, it is renamed here.)

Note: Soil #1: 9 ppm total PAHs; Soil #2: 377 ppm total PAHs.

## TABLE 6

## BENZO(a)PYRENE BIOAVAILABILITY FROM SOILS[a]
## (GOON et al., 1991)

| SOIL AGING | SANDY SOIL | CLAY-BASED SOIL |
|---|---|---|
| 1 day | 66.9% | 48.8% |
| 1 week | 70.4% | 52.1% |
| 1 month | 67.7% | 58.5% |
| 6 months | 54.3% | 38.5% |
| 1 year | 62.2% | 38.6% |
| Average- 6 mo. & 1 year | 58.3% | 38.6% |

a (Area under the blood radioactivity curve)$_{soil}$ / (Area under blood radioactivity curve)$_{solution}$

## TABLE 7

## ANIMAL WEIGHTS DURING GOON et al. (1990)

| TREATMENT GROUP | SEX | NONFASTED WEIGHT (g) | FASTED WEIGHT (g) | WEIGHT AT DAY 7 (g) |
|---|---|---|---|---|
| Solution | Males | 221 +/- 9 | 218 +/- 12 | 237 +/- 24 |
| | Females | 175 +/- 3 | 165 +/- 4 | 179 +/- 3 |
| Rodent Chow | Males | 226 +/- 7 | 222 +/- 2 | 255 +/- 5 |
| | Females | 173 +/- 6 | 165 +/- 2 | 193 +/- 4 |
| Sandy Soil | Males | 222 +/- 6 | 216 +/- 6 | 228 +/- 19 |
| | Females | 180 +/- 4 | 167 +/- 4 | 190 +/- 4 |
| Clay-Based Soil | Males | 230 +/- 5 | 221 +/- 8 | 251 +/- 10 |
| | Females | 172 +/- 6 | 162 +/- 6 | 188 +/- 4 |

**TABLE 8**

**WEIGHT GAIN DURING GOON et al. (1990)**

| TREATMENT GROUP | SEX | WEIGHT GAIN DURING 7 DAY PERIOD (g) | % WEIGHT GAIN DURING 7 DAY PERIOD |
|---|---|---|---|
| Solution | Males | 19 | 9% |
| | Females | 14 | 8% |
| Rodent Chow | Males | 33 | 15% |
| | Females | 28 | 17% |
| Sandy Soil | Males | 12 | 6% |
| | Females | 23 | 14% |
| Clay-Based Soil | Males | 30 | 14% |
| | Females | 26 | 16% |

## TABLE 9

## ANIMAL WEIGHTS DURING GOON et al. (1991)

| AGING PERIOD | TREATMENT GROUP | NONFASTED WEIGHT (g) | FASTED WEIGHT (g) | WEIGHT AT DAY 7 (g) |
|---|---|---|---|---|
| 1 DAY | Solution | 238 +/- 3 | 219 +/- 3 | 255 +/- 4 |
| 1 DAY | Clay-Based Soil | 245 +/- 4 | 235 +/- 4 | 252 +/- 11 |
| 1DAY | Sandy Soil | 256 +/- 6 | 239 +/- 6 | 266 +/- 11 |
| 1 WEEK | Clay-Based Soil | 222 +/- 3 | 217 +/- 3 | 243 +/- 10 |
| 1WEEK | Sandy Soil | 223 +/- 4 | 216 +/- 3 | 241 +/- 9 |
| 1 MONTH | Clay-Based Soil | 243 +/- 8 | 220 +/- 5 | 254 +/- 7 |
| 1 MONTH | Sandy Soil | 241 +/- 5 | 219 +/- 4 | 268 +/- 4 |
| 6 MONTHS | Clay-Based Soil | 238 +/- 3 | 211 +/- 3 | 263 +/- 5 |
| 6 MONTHS | Sandy Soil | 244 +/- 3 | 217 +/- 4 | 263 +/- 5 |
| 1 YEAR | Clay-Based Soil | 242 +/- 5 | 214 +/- 4 | 259 +/- 7 |
| 1 YEAR | Sandy Soil | 244 +/- 5 | 214 +/- 4 | 258 +/- 7 |
| | | | | |

## TABLE 10

## ORAL-SOIL AAFS FOR PAHS

| Oral-Soil AAF | Notes | Source |
|---|---|---|
| 0.07 | CD-1 mice, MGP soil, 0.71-0.85 mm | Rozett *et al.* (1996) |
| 0.07 | CD-1 mice, MGP soil, 0.6-0.71 mm | Rozett *et al.* (1996) |
| 0.08 | CD-1 mice, MGP soil, 0.5-0.6 mm | Rozett *et al.* (1996) |
| 0.09 | CD-1 mice, MGP soil, 0.15-0.3 mm | Rozett *et al.* (1996) |
| 0.11 | $B_6C_3F_1$ mice, MGP soil | Weyand *et al.* (1996) |
| 0.28 | CD-1 mice, MGP soil, <1 mm | Rozett *et al.* (1996) |
| 0.32 | CD-1 mice, MGP soil, 0.85-1 mm | Rozett *et al.* (1996) |
| 0.36 | $B_6C_3F_1$ mice, MGP soil | Weyand *et al.* (1996) |
| 0.37 | rats, clay-based soil | Goon *et al.* (1991) |
| 0.40 | CD-1 mice, MGP soil, 0.3-0.5 mm | Rozett *et al.* (1996) |
| 0.57 | rats, sandy soil | Goon *et al.* (1991) |
| 0.76 | CD-1 mice, MGP soil, <0.15 mm | Rozett *et al.* (1996) |

## TABLE 11

## DERMAL ABSORPTION OF BENZO(a)PYRENE FROM SOIL IN THE RAT YANG, *ET AL.* (1989)

| Time Point | *In Vivo* Results | *In Vitro* Results |
|---|---|---|
| 24 Hours[1] | 1.1% (0.3)[1,2] | 1.5%[4] |
| 48 Hours[1] | 3.7% (0.8)[1,2] | 3.5%[4] |
| 72 Hours[1] | 5.8% (1.0)[1,2] | 5.5%[4] |
| 96 Hours[3] | 9.2% (1.2)[1,3] | 8.4%[4] |

[1]Values shown for 48-96 hours are cumulative. Results are the mean for five rats (standard error).
[2] Urine plus feces
[3] Urine plus feces plus tissues.
[4] See Figure 1 of Yang, *et al.* (1989)

## TABLE 12

## DERMAL ABSORPTION OF BENZO(a)PYRENE FROM SOIL WESTER, *ET AL.* (1990)

| Sample | Monkey Skin | Human Skin |
|---|---|---|
| 1 | 13.1%[1] | 1.01%[3] |
| 2 | 10.8%[1] | 1.52%[3] |
| 3 | 18.0%[1] | 0.61%[3] |
| 4 | 11.0%[1] | 2.21%[3] |
| 5 | NA | 0.31%[3] |
| 6 | NA | 3.01%[3] |
| Mean +/- SD | 13.2% +/- 3.4%[2] | 1.45% +/- 1.02% [2] |

[1]Percentage of applied dose absorbed = ($^{14}C$ urinary excretion for seven days following 24 hour topical application) /($^{14}C$ urinary excretion following intravenous administration) x 100

[2] Mean +/- Standard Deviation

[3] Fraction of applied dose in the skin plus fraction in receptor fluid.

## REFERENCES

Bartosek, I., Guitani, A., Modica, R., Fiume, M., and Urso, R. 1984. Comparative Kinetics of Oral Benz(a)Anthracene, Chrysene, and Triphenylene in Rats: A Study with Hydrocarbon Mixtures. Toxicology Letters 23:333-339.

Bonaccorsi, A., di Domenico, A., Fanelli, R., Merli, F., Motta, R., Vanzati, R., and Zapponi, G.A. 1984. The Influence of Soil Particle Adsorption on 2,3,7,8-Tetrachlorodibenzo-p-dioxin Biological Uptake in the Rabbit. *Arch. Toxicol. Suppl.* 7:431-434.

Bowes, S.G. and Renwick, A.G. 1986. The hepatic metabolism and biliary excretion of benzo[a]pyrene in guinea-pigs fed normal, high-fat or high-cholesterol diets. *Xenobiotica* 16:531-542.

Brune, H., Deutsch-Wenzel, R.P., Habs, M., Ivankovic, S., and Schmahl, D. 1981. Investigation of the tumorigenic response to benzo(a)pyrene in aqueous caffeine solution applied orally to Sprague-Dawley rats. *J. Cancer Res. Clin. Oncol.* 102(2):153-157.

Burmaster, D.E. 1996. Using Beta Distributions Efficiently in a Probabilistic Exposure Assessment. Submitted to *Risk Analysis.*

Chang, L.H. 1943. The fecal excretion of Polycyclic Hydrocarbons Following Their Administration to the Rat. *J. Biol. Chem.* 151:93-99.

Chipman, J.K., Hirom, P.C., G.S. Frost, G.S., and Millburn, P. 1981a. The Biliary Excretion and Enterohepatic Circulation of Benzo(a)pyrene and its Metabolites in the Rat. *Biochemical Pharmacology* 30:937-944.

Chipman, J.K., Frost, G.S., Hirom, P.C., and Millburn, P. 1981b. Biliary excretion, systemic availability, and reactivity of metabolites following intraportal infusion of [3H]benzo[a]pyrene in the rat. *Carcinogenesis* 2:741-745.

Clinton, S.K. and Visek, W.J. 1988. Wheat Bran and the Induction of Intestinal Benzo(a) pyrene Hydroxylase by Dietary Benzo(a)pyrene. *J. Nutr.* 119:395-402.

Clonfero, E., Zordan, M., Cottica, D., Venier, P., Pozzoli, L., Cardin, E.L., Sarto, F., and Levis, A.G. 1986. Mutagenic activity and polycyclic aromatic hydrocarbon levels in urine of humans exposed to therapeutical coal tar. *Carcinogenesis* 7(5): 819-823.

Dankovic, D.A., Wright, C.W., Zangar, R.C., and Springer, D.L. 1989. Complex Mixture Effects on the Dermal Absorption of Benzo[a]pyrene and other Polycyclic Aromatic Hydrocarbons from Mouse Skin. *Journal of Applied Toxicology*, 9(4): 239-244.

Edwards, C.A., Beck, S.D., and Lichtenstein, E.P. 1957. Bioassay of aldrin and lindane in soil. *J. Econ. Entomol.* 50:622-626. (cited in GRI, 1995)

EPA. 1992a. Guidelines for Exposure Assessment. (Final). 57 *Federal Register* 22888-22938.

EPA. 1992b. Dermal Exposure Assessment Principles and Applications. Office of Research and Development. Washington, D.C. EPA 600/8-91/011B.

EPA. 1993. Provisional Guidance for Quantitative Risk Assessment of Polycyclic Aromatic Hydrocarbons. Office of Research and Development, Washington, D.C. EPA/600/R-93/089

Flescher, J.W. and Syndor, K.L. 1960. Distribution and Excretion of Radioactivity in Rats after Oral Administration of H3-3-Methylcholanthrene. *Proc. Soc. Expl. Biol. Med.* 104:776-779.

Foth, H., Kahl, R. and Hahl, G.F. 1988. Pharmacokinetics of low doses of benzo[a]pyrene in the rat. *Fd. Chem. Toxic.* 26:45-51.

Gas Research Institute. 1990. Manufactured Gas Plant Update:. A Publication of the Gas Research Institute (GRI) Environment & Safety Research Department.

Gas Research Institute. 1995. Environmentally Acceptable Endpoints in Soil: Risk-Based Approach to Contaminated Site Management Based on Availability of Chemicals in Soil. April, 1995. Draft Report.

Grimmer, G., Brune, H., Dettbarn, G., Heinrich, U., Jacob, J., Mohtashamipur, E., Orpoth, K., Pott, F., and Wenzel-Hartung, R. 1988. Urinary and faecal excretion of chrysene and chrysene metabolites by rats after oral, intreperitoneal, intratracheal or intrapulmonary application. *Arch. Toxicol.* 62:401-405.

Gulliver, W.P., Kutty, K.P., Laher, J.M., and Barrowman, J.A. 1983. In Vitro Interaction of 7,12-dimethylbenz[a]anthracene and its Biliary Metabolites with Dietary Fibers. *J. Nat. Cancer Inst.* 71:207-210.

Goon, D., N.S. Hatoum, J.D. Jernigan, S.L. Schmidt, and P.J. Garvin. 1990. Pharmacokinetics and Oral Bioavailability of Soil-Adsorbed Benzo(a)pyrene (B(a)P in Rats. *Toxicologist* 10:218.

Goon, D., N.S. Hatoum, M.J. Klan, J.D. Jerniganm and R.G. Farmer. 1991. Oral Bioavailability of "Aged" Soil-Adsorbed Benzo[a]pyrene (BaP) in Rats. *Toxicologist* 11:1356.

Goon, D. and Don Burnette. 1996. Personal Communication.

Hecht, S.S., Grabowski, W. and Groth, K. 1979. Analysis of Feces for Benzo(a)pyrene After Consumption of Charcoal-Broiled Beef by Rats and Humans. *Cosmet. Toxicol.* 17:223-227.

Hatzinger, P.B. and Alexander, M. 1995. Effect of aging of chemicals in soil on their biodegradability and extractability. *Environ. Sci. Technol.* 29:537-545.

Huq, A.S., Ho, N.F.H., Husari, N., Flynn, G.L., Jetzer, W.E., and Condie, L. 1986. Permeation of Water Contaminative Phenols Through Hairless Mouse Skin. *Arch. Environ. Contam. Toxicol.* 15:557-566.

Kao, J., Patterson, F.K., and Hall, J. 1985. Skin Penetration and Metabolism of Topically Applied Chemicals in Six Mammalian Species, Including Man: An In Vitro Study With Benzo(a)pyrene and Testosterone. *Toxicol. Appl. Pharmacol.* 81:502-516.

Kawamura, J., Kamata, E., Ogawa, Y., Kaneko, T., Uchiyama, S., and Saito, Y. 1988. The Effect of Various Foods on the Intestinal Absorption of Benzo(a)pyrene in Rats. *J. Food Hyg. Soc. Japan* (29):21-25.

McConnell, E., Lucier, G., Rumbaugh, R., Albro, P., Harvan, D., Hass, J., and Harris, M. 1984. Dioxin in soil: bioavailability after ingestion by rats and guinea pigs. *Science* 223: 1077-1079.

Mirvish, S.S., Ghadirian, P., Wallcave, L., Raha, C., Bronczyk, S., and Sams, J.P. 1981. Effect of Diet on Fecal Excretion and Gastrointestinal Tract Distribution of Unmetabolized Benzo(a)pyrene and 3-Methylcholanthrene When These Compounds Are Administered Orally to Hamsters. *Cancer Research* 41:2289-2293.

Neal, J. and Rigdon, R.H. 1967. Gastric Tumors in Mice Fed Benzo(a)pyrene: A Quantitative Study. *Texas Rep. Biol. Med.* 25:553.

Ng, K.M.E., Chu, I., Bronaugh, R.L., Franklin, C.A., and Somers, D.A. 1992. Percutaneous Absorption and Metabolism of Pyrene, Benzo[a]pyrene, and Di(2-ethylhexyl) Phthalate: Comparison of *in Vitro* and *in Vivo* Results in the Hairless Guinea Pig. *Toxicology and Applied Pharmacology* 115: 216-223.

Peterson, J. R., Adams, R.S. and Cutkomp, L.K. 1971. Soil properties influencing DDT bioactivity. *Soil Sci. Soc. Am. Proc.* 35:72-78 (cited in GRI, 1995).

Poiger, H. and Schlatter, Ch. 1980. Influence of Solvents and Adsorbents on Dermal and Intestinal Absorption of TCDD. *Food Cosmet. Toxicol.* 18:477-481.

Rabache, M.C. Billaud and Adrian, J. 1985. Evolution du benzo(a)pyrene dans le tube digestif. 1. Sa disparition. Internat. *J. Vit. Nutr. Res.* 55:113-117.

Rees, E.D., Mandelstam, P., Lowry, J.Q., and Lipscomb, H. 1971. A study of the mechanism of intestinal absorption of benzo(a)pyrene. *Biochemica et Biophysica Acta* 225:96-107.

Roberts, M.S., Anderson, R.A., and Swarbrick, J. 1977. Permeability of human epidermis to phenolic compounds. *Pharm. Pharmac.* 29:677-683.

Rozett, K., R. Singh, T. Roy, W. Neal, and E.H. Weyand. 1996. *Bioavailability of Chemical Components of Soil Contaminated with Maunfactured Gas Plant Residue.* Fundamental and Applied Toxicology 30:1, Part 2.

Sanders, C.L., Skinner, C., Gelman, R.A. 1984. Percutaneous Absorption of [7, $10^{-14}$ C] Benzo[a]Pyrene and [7,12-14C] Dimethylbenz[a]Anthracene in Mice. *Environmental Research* 33: 353-360.

Sato, T., Ose, Y., Nagase, H., and Hayase, K. 1987. Mechanism of the desmutagenic effect of humic acid. *Mutation Research* 176:199-204.

Schoket, B., Horkay, I., Kosa, A., Paldeak, L., Hewer, A., Grover, P.L., and Philips, D.H. 1990. Formation of DNA Adducts in the Skin of Psoriasis Patients, in Human Skin in Organ Culture, and in Mouse Skin and Lung Following Topical Application of Coal-Tar and Juniper Tar. *Journal of Investigative Dermatology.* pp. 241-246.

Singh, R., Tucek, M., Maxa, K., Tenglerova, T., and Weyand, E.H. 1995. A Rapid and Simple Method for the Analysis of 1-hydroxypyrene glucuronide: A Potential Biomarker for Polycyclic Aromatic Hydrocarbon Exposure. *Carcinogenesis* 16:2909-2915.

Storer, J.S., DeLeon, I., Millikan, L.E., Laseter, J.L., and Griffing, C. 1984. Human Absorption of Crude Coal Tar Products. *Arch. Dermatol.* 120: 874-877.

Storm, J.E., Collier, S.W., Stewart, R.F., and Bronaugh, R.L. 1990. Metabolism of Xenobiotics during Percutaneous Penetration: Role of Absorption Rate and Cutaneous Enzyme Activity. *Fundamental and Applied Toxicology* 15: 132-141.

Umbreit, T.H., Hesse, E.J., and Gallo, M.A. 1986. Bioavailability of Dioxin in Soil from a 2,4,5-T Manufacturing Site. *Science*. 232:497-499.

Van den Berg, M., Olie, K., Hutzinger, O. 1983. Uptake and selective retention in rats of orally administered chlorinated dioxins and dibenzo-furans from fly-ash and fly-ash extract. *Chemosphere.* 12(4/5): 537-544.

VanRooij, J.G.M., Vinke, E., De Lange, J., Bruijnzeel, P.L.B., Bodelier-Bade, M.M., Noordhoek, J., and Jongeneelen, F.J. 1995. Dermal Absorption of Polycyclic Aromatic Hydrocarbons in the Blood-perfused Pig Ear. *Journal of Applied Toxicology* 15(3):193-200.

Weissenfels, W.D., Klewer, H.J., and Langhoff, J. 1992. Adsorption of Polycyclic Aromatic Hydrocarbons (PAHs) by Soil Particles; Influence on Biodegradability and Biotoxicity. *Appl. Microbiol. Biotechnol.* 36:689-696.

Wester, R.D., Maibach, H.I., Bucks, D.A., Sedik, L., Melendres, J., Liao, C., and DiZio, S. 1990. Percutaneous Absorption of [$^{14}$C]DDT and [$^{14}$C] Benzo(a)pyrene from Soil. *Fundam. Appl. Toxicol.* 15:510-516.

Weyand, E.H. and Wu, Y. 1995. Covalent binding of polycyclic aromatic hydrocarbon components of manufactured gas plant residue to mouse lung and forestomach DNA. *Chem. Res. Toxicol.* 8:955-962.

Weyand, E.H., Rozett, K., Koganit, A., and Singh, R. 1996. Effect of Soil on the Genotoxicity of Manufactured Gas Plant Residue. *Fundamental and Applied Toxicology* 30:1, Part 2.

Withey, J.R., Law, F.C.P., and Endrenyi, L. 1991. Pharmacokinetics and Bioavailability of Pyrene in the Rat. *J. Toxicol. Environ. Health* 32:429-447.

Yang, J.L., Roy, T.A., and Mackerer, C.R. 1986a. Percutaneous Absorption of Anthracene in the Rat: Comparison of *in vivo* and *in vitro* results. *Toxicology and Industrial Health*, 2(1): 79-84.

Yang, J.L., Roy, T.A., and Mackerer, C.R. 1986b. Percutaneous Absorption of Benzo[a]Pyrene in the Rat: Comparison of *in vivo* and *in vitro* results. *Toxicology and Industrial Health*, Vol. 2, No. 4.

Yang, J.J., Roy, T.A., Krueger, A.J., Neil, W., and Mackerer, D.R. 1989. In Vitro and In Vivo Percutaneous Absorption of Benzo(a)pyrene From Petroleum Crude-Fortified Soil in the Rat. *Bull. Environ. Contam. Toxicol.* 43:207-214.

# Appendix H

# Uncertainty Analysis of Risk Estimates

25 October 1996

Cox Associates
503 Franklin Street
Denver, CO 80218

303-541-6043

Appendix H

# UNCERTAINTY ANALYSIS OF RISK ESTIMATES

This appendix examines uncertainties in the exposure and health effects data that are relevant for assessing potential human health risks associated with exposure to contaminants originating from the industrial property. Many of the key quantities considered in the risk assessment are highly uncertain. For example, the following factors have not been estimated with high precision or confidence:

(a) The spatial distribution and extent of contamination from the industrial property in various directions is not well known. Instead, it must be estimated from soil sample data. (See discussion in Appendix B).

(b) The fraction of PAHs found at any specific location that arise from the industrial property is uncertain. The problem of distinguishing between Celotex-related and "background" (meaning non-Celotex-related) contamination arises, since the same contaminants and approximate composition of PAHs found near the industrial property are also found at distances remote enough to make association with the industrial property implausible.

(c) The magnitudes and frequencies of individual exposures depend on individual behaviors and on details of the yards (e.g., extent of vegetative cover as opposed to rock and debris cover) that have not been quantified. Hence, the actual magnitude of individual exposures is uncertain. Drive-by inspection of yards in the vicinity of the industrial property suggests that they are dissimilar in many respects (e.g., more rubble, less accessible soil useful for gardening or recreation) compared to locations further from the industrial property. How these local characteristics affect individual behaviors and exposures has not been estimated. Similarly, local demographic characteristics (e.g., the ages, occupations, recreational patterns, etc.) of neighbors of the industrial property have not been examined. Yet, these characteristics may affect the magnitudes and frequencies of individual exposures to yard soils.

(d) The amounts of internal doses of reactive, potentially carcinogenic PAH metabolites formed in humans at the exposure levels in question are not known. In particular, the relative amounts of internal doses formed in humans compared to the amounts formed in animals under the experimental conditions used to establish the carcinogenicity of PAHs such as B(a)P are not known.

(e) The cancer potency of PAHs, including B(a)P, at the concentrations found near the industrial property is not known. Specifically, the relation between carcinogenic potency of B(a)P at the high doses used in animal carcinogenicity experiments and its potency at the much lower levels found in the soil samples examined in this study is not known. In addition, the potency of the PAH mixtures found in the soil samples is uncertain.

These uncertainties create a challenge for fair, efficient, health-protective risk management. The actual human health risks posed by the industrial property are not known. They would be costly to quantify with high precision and confidence, since doing so would require resolving each of these sources of uncertainty. Yet, it is desirable to avoid the two types of risk management errors most likely to occur in this case: failure to adequately reduce Celotex-related exposures, and failure to limit reductions to those that significantly reduce actual human health risks. The purpose of the analyses reported in this appendix and the next one is to reduce the probabilities of both types of errors by introducing relevant information and findings from recently completed data analyses and literature reviews. A suggested approach to risk management decision-making in the presence of the uncertainties just listed is offered after some relevant facts, data, and statistical results have been summarized.

Appendix B focuses on the first issue -- uncertainty about the spatial extent of contamination from the industrial property. This issue can be addressed without considering risk assessment questions and data: it rests solely on statistical analysis of the soil sample data collected so far. Analysis of these data reveals the maximum probable geographic extent of effects from the property, and thus provides a basis for bounding the geographic scope of the risk assessment without regard for risk magnitudes. This appendix presents the remaining sources of uncertainty and their implications for risk management.

# ESTIMATING THE FRACTION OF PAH EXPOSURES AND CARCINOGEN EXPOSURES DUE TO THE CELOTEX PROPERTY

## CONCEPTS FOR ASSIGNED SHARE CALCULATIONS

Cancer risks due to environmental chemical carcinogens are typically small compared to cancer risks from all sources. At most, a few percent of total avoidable cancer risk is typically attributable to environmental insults (Doll and Peto, 1981). This raises the following key question: *How much of the excess risk of cancer created by PAHs in the soil at properties examined in this study is due to the industrial property,* as opposed to other sources of PAHs? The calculations of the previous appendix can help to answer this question.

The scientific, demographic, and behavioral uncertainties previously catalogued make it impossible to determine absolute risks with high precision and confidence. However, the attributable risk of excess cancers due to the industrial property may be estimated by the ratio of Celotex-related exposure to total exposure, with the exposures from different sources being weighted by their relative potencies when these can be estimated.

The probability that an excess cancer is attributable to soil PAHs from Celotex operations can be estimated as the product of the following four factors, each of which addresses an aspect of the relative contribution to total carcinogenic exposure made by different sources:

1. PAH from Celotex-contaminants in soil / PAH from all contaminants in soil.

2. PAH from all soil / PAH from all sources (cooked foods, cigarettes, diesel exhaust emissions, coal-fired power plant emissions, etc.)

3. Exposure to PAH from all sources / Exposure to all carcinogens from all environmental sources, weighted by their relative potencies.

(Note: Other sources of environmental carcinogens identified by the EPA include diesel exhaust, radon, cigarette smoke, benzene, MTBE, vinyl chloride, and so forth. This background of exposures to carcinogens has not been quantified specifically for neighbors of the industrial property.)

4. Exposure to carcinogens from all environmental sources / Exposure to carcinogens from all sources (including diet and lifestyle).

In the absence of more specific information about the values of these factors, the relative contribution of Celotex-related contamination to the relative risk of cancer can be estimated by taking a plausible upper bound on each of the four factors and using their product as an upper-bound estimate for the conditional probability that Celotex contamination would be responsible for a cancer, given that a cancer occurs.

Table 1 contains background information on B(a)P that may be helpful in estimating plausible upper bounds for the first three of these factors. For example, to quantify the fraction of soil PAHs that are due to the industrial property, one must consider other, competing sources of soil PAHs, including

(i) Soil PAHs from smoke or soot (e.g., from wood fires, outdoor cooking or charcoal grilling, pollution from coal-fired power plants, etc.)

(ii) Soil PAHs from automobile emissions and other gasoline combustion (e.g., lawn mowers, other gasoline-powered or diesel-powered equipment).

(iii) Soil PAHs from asphalt roads (e.g., carried in water runoff).

Although the fractions of soil PAHs due to each of these sources have not been quantitatively modeled for locations near the industrial property, it would be unrealistic to completely ignore them. One possibility is to represent the fraction of soil PAHs due to emissions from the industrial property as an unknown quantity, uniformly distributed between 0 and some plausible upper bound. To be conservative (i.e., tending to maximize the risk attributed to the industrial property), the upper bound might be taken to be 100%.

TABLE 1: Excerpts from Agency for Toxic Substances and Disease Registry (ATSDR) Public Health Statement for B(a)P, May, 1990

*What is benzo[a]pyrene?*

Benzo[a]pyrene (B[a]P) is one of the polycyclic aromatic hydrocarbon (PAH) compounds. Because it is formed when gasoline, garbage, or any animal or plant material burns, it is usually found in smoke and soot. This chemical combines with dust particles in the air and is carried into water and soil and onto crops. Benzo[a]pyrene is found in the coal tar pitch that industry uses to join electrical parts together. It is also found in creosote, a chemical used to preserve wood.

*How might I be exposed to benzo[a]pyrene?*

People may be exposed to B[a]P from environmental sources such as air, water, and soil and from cigarette smoke and cooked food. Workers who handle or are involved in the manufacture of PAH-containing materials may also be exposed to B[a]P. Typically, exposure for workers and the general population is not to B[a]P alone but to a mixture of similar chemicals.

The soil near areas where coal, wood, or other products have been burned is another source of exposure. Exposure to B[a]P and other PAHs may also occur through skin contact with products that contain PAHs such as creosote-treated wood, asphalt roads, or coal tar.

People may be exposed to B[a]P by drinking water from the drinking water supplies in the United States that have been found to contain low levels of the chemical. Foods grown in contaminated soil or air may contain B[a]P. Cooking food at high temperatures, as occurs during charcoal-grilling or charring, can increase the amount of B[a]P in the food. Benzo[a]pyrene has been found in cereals, vegetables, fruits, meats, beverages, chewing tobacco, and in cigarette smoke.

The greatest exposure to B[a]P is likely to take place in the workplace. People who work in coal tar-production plants; coking plants; asphalt-production plants; coal-gasification sites; smoke houses; municipal trash incinerators; and facilities that burn wood, coal, or oil may be exposed to B[a]P in the workplace air. Benzo[a]pyrene may also be found in areas where high-temperature food fryers and broilers are used.

The general population may be exposed to dust, soil, and other particles that contain B[a]P. The largest

sources of B[a]P in the air are open burning and home heating with wood and coal. Factories that produce coal tar also contribute small amounts of B[a]P to the air. People may come in contact with B[a]P from soil on or near hazardous waste sites, such as former gas-manufacturing sites or abandoned wood-treatment plants that used creosote.

*How does benzo[a]pyrene get into my body?*

The most common way B[a]P enters the body is through the lungs when a person breathes in air or smoke containing it. It also enters the body through the digestive system when substances containing it are swallowed. Although B[a]P does not normally enter the body through the skin, small amounts could enter if contact occurs with soil that contains high levels of B[a]P (for example, near a hazardous waste site) or if contact is made with heavy oils containing B[a]P.

*What levels of exposure have resulted in harmful health effects?*

No information has been found about specific levels of B[a]P that have caused harmful effects in people after breathing, swallowing, or touching the substance.

Similarly, the fraction of total PAH exposure that is due to soil PAHs can only be calculated or estimated by considering competing sources of exposure, such as cooked foods, cereal, vegetables, meat, and fruits, home heating with wood or coal, drinking water, secondary cigarette smoke, and so forth (see Table 1). Given all these sources of PAH exposure, the fraction due to PAHs in soil may be quite small, especially if the different sources are weighted to reflect relative bioavailabilities of the PAHs from different sources (with PAHs in food probably being more readily available than PAHs in soil particles, for example). A subjective estimate of a plausible upper bound on the fraction of PAH exposure due to soil might be 10%, although the true number could be much lower, depending on details of cigarette smoking, heating fuels, consumption of cooked foods, and so forth that have not yet been provided for neighbors of the industrial property. To model the uncertainty about this fraction, a uniform distribution between 0 and 0.1 might be assumed.

The fraction of total environmental carcinogen exposure that is attributable to PAHs is even more difficult to estimate in the absence of specific information about individual behaviors and exposures. The Agency for Toxic Substances and Disease Registry (ATSDR) that provided the information in Table 1 also maintains a ranked list of the top 20 hazardous substances, on which B(a)P appears as the eighth one (after lead, arsenic, metallic mercury, vinyl chloride, benzene, PCBs, cadmium, and before chloroform, various pesticides, and other chemicals). The ATSDR links each of the chemicals ranked above B(a)P to potential human cancers, with the sole exception of metallic mercury. A reasonable assumption might be that the carcinogenic burden from PAHs accounts for no more than 5% of the total carcinogenic burden imposed by all hazardous substances (since there are many in the top 20 alone that may pose an equal or greater threat).

Rather than making additional speculative assumptions about the probability distributions of the four relative risk factors, one could use deterministic upper bounds to calculate an upper bound on the excess cancer risk attributable to the industrial property. For example, if the upper bounds used are as follows:

1. Celotex contribution to PAHs in soil < 100%

2. Soil contribution to total environmental PAH exposure < 10% (based on prevalence of secondary smoke, engine exhaust emissions, air pollution, and other sources of environmental PAHs)

3. Environmental PAH contribution to total environmental carcinogen exposure < 5% (based on prevalence of non-PAH carcinogens such as benzene, radon gas , pesticides, drinking water carcinogens and other sources of environmental carcinogens)

4. Environmental contribution to total carcinogen exposure < 2% (based on Doll and Peto estimates from cancer epidemiology)

then the conditional probability, or share in risk, for excess cancer risk attributable to the industrial property would not be expected to exceed the following:

Example upper bound for risk attributable to the industrial property <u>if</u> a cancer occurs = (100%)(10%)(5%)(2%) = 0.0001.

The risk due to Celotex contamination may be smaller, although it cannot be larger (unless the estimated upper bounds are too small). For example, if the effective cancer potency of soil PAHs at the concentrations found in yards located near the industrial property were very small or zero, then the absolute risk due to these contaminants would, correspondingly, be very small or zero. The value of 0.0001 is based solely on estimated relative contributions to carcinogenic <u>exposures</u>, rather than on the abilities of such exposures to cause cancer.

Instead of making a deterministic upper-bound calculation, it may be preferable to take a more informative probabilistic approach that better expresses uncertainties about the various risk factors. For example, suppose that the following probability assumptions are made:

1. Celotex contribution to PAHs in soil is uniformly distributed between 0 and 100%.

2. Soil contribution to total environmental PAH exposure is uniformly distributed between 0 and 10%.

3. Environmental PAH contribution to total environmental carcinogen exposure is uniformly distributed between 0 and 5%.

4. Environmental contribution to total carcinogen exposure is uniformly distributed between 0 and 2%.

Then Monte-Carlo uncertainty analysis shows that the expected excess risk (or relative exposure) attributable to the industrial property is about 6E-06, i.e., 6 parts in a million, with a 95% upper uncertainty bound of about 30E-06. These numbers are again based on estimated relative contribution of Celotex-related soil PAHs to total carcinogenic burden. The actual risk due to the Celotex-related PAHs may be smaller if the potencies of carcinogens are taken into account. Specifically, as we shall next show, the carcinogenic potency of soil PAHs in humans may be quite low. Thus, the calculations in this section should be interpreted as attempting to establish plausible bounds on the uncertainty about the contribution of the industrial property to cancer risk.

# ESTIMATING THE INTERNAL DOSES OF CARCINOGENS FORMED FOLLOWING EXPOSURE TO SOIL PAHs

## CONCEPTS FOR INTERNAL DOSE CALCULATIONS

The logic of the preceding calculations is that if no more than x% of a set of identical cancer-causing molecules acting on a person come from a particular source, then, in the absence of more specific information, no more than x% of the cancer risk experienced by the person should be attributed to that source. Such calculations can be very useful when there is a lot of uncertainty about absolute exposure magnitudes and cancer potencies, because the "percentage of molecules" perspective does not depend on the absolute number of molecules or on their ability to cause cancer. To go further and seek to quantify absolute risk, as defined by the expected number of excess tumors created by exposure to a source (such as soil PAHs originating at the industrial property), it is necessary to make some speculative assumptions. As noted by the Federal EPA (IRIS data base), "Human data specifically linking benzo(a)pyrene (BAP) to a carcinogenic effect are lacking." Instead, animal data have been used to estimate the potential carcinogenicity of B(a)P in humans. But this extrapolation raises several additional uncertainties. For example:

1. The routes of dose administration used in animal studies (e.g., implantation in the stomach wall, injection, ingestion, forced inhalation, intratracheal instillation) are not representative of realistic exposure conditions. Cancers that occur under such extreme conditions of administration may not occur under more realistic conditions, as has been demonstrated experimentally in animals (Collins et al., 1991, p. 171). Moreover, neither the inhalation route nor the forced ingestion routes studied in animal experiments are necessarily relevant to the exposure pathways (especially, dermal and ingestion) experienced by humans.

2. The extrapolation of effects from rodents to humans is speculative. To make it more credible, the ways in which humans and animals metabolize and eliminate PAH (and, specifically, B(a)P) doses must be compared.

3. The interpolation of tumorigenic response rates between high doses and low (or zero) doses is questionable. The best procedures for interpolating between control group tumor rates and dose group tumor rates will depend on the biological processes involved.

Current regulatory risk estimates either ignore these uncertainties or introduce simple "default" assumptions to fill in the gaps in scientific knowledge. The remainder of this section reexamines the inter-species extrapolation and high-dose to low-dose interpolation questions using data specific to B(a)P, which are more relevant in this case than the default assumptions.

A key concept of modern, "biologically-based" risk assessment (BBRA) is that administered doses affect cancer rates and cancer risks only through internal doses, e.g., through the quantity of reactive carcinogenic metabolites formed by metabolic activation of the administered PAH. Therefore, the following two questions become central in assessing the implications of animal experiment data for human cancer risks:

*Q1. How does the probability of tumor, in a human or in an animal, depend on the quantity of internal dose received?*

*Q2: How does the internal dose received depend on the dose administered* (or on "exposure", in the case of humans)?

The answers to these two questions determine the absolute risk associated with a given exposure profile or administered dose.

## DIFFERENCES IN INTERNAL DOSES ACROSS SPECIES BASED ON COMPARISONS OF ENZYME ACTIVITY LEVELS

In the situation of chronic, low-level exposures most relevant for the Celotex case, standard pharmacokinetic and metabolic models imply that the average internal dose of carcinogenic metabolites (whose identities may or may not be known) reaching the (perhaps unknown) target organs and cell populations per unit time will be proportional to the average administered dose

per unit time (Cox, 1995). However, the steady-state ratio of internal dose to administered dose may be different in different species. A standard "default" assumption made in many regulatory risk assessments, including those for B(a)P, is that the ratio of biologically effective internal dose -- meaning the dose of carcinogenic metabolites acting on cell populations to cause cancer or increase cancer risk -- to administered dose depends on the body weight of the exposed species. (This includes allometric scaling based on surface areas, since surface areas are determined from body weights.) For specific chemicals such as B(a)P, it is possible to replace this generic default assumption with more specific and relevant information. For example, available evidence on the biochemistry of metabolic activation and detoxification off B(a)P in different species may be used to refine the estimated ratio of internal dose to administered dose in different species. This is often done by examining the enzymes involved in metabolic activation and detoxification and comparing the activity intensities of these enzymes across species.

It is generally accepted that many carcinogenic PAHs, including B(a)P, are metabolized to their carcinogenic forms by monooxygenases (specifically, the P450 enzyme superfamily, which catalyzes single-electron oxidation of PAHs and binding to DNA, as well as playing a role in subsequent detoxification of diol intermediates) (Cavalieri and Rogan, 1992). Recent evidence also shows that protective enzymes (glutathione) that shield cells from oxidative DNA damage probably play a role in PAH carcinogenicity in both humans (Grinberg-Funes et al., 1994) and other species (Kirby et al., 1995). The highest carcinogenic risks from PAH exposures are likely to occur under exposure conditions that cause depletion of glutathione (GST and/or GSH) reserves and that lead to high levels of metabolically activated PAH oxidized metabolites. Conversely, if exposure levels are low enough so that glutathione reserves are adequate and carcinogenic metabolites of PAHs are removed or detoxified before they can bind to DNA or other target macromolecules and cause damage, then cancer risk is likely to be relatively low.

These elementary biochemical observations suggest that the ratio of monooxygenase (P450) to glutathione resources in different species may provide a useful qualitative guide to species susceptibility to PAH-induced carcinogenesis. Quantitative data for making comparisons across species are

suggestive at best, but may be obtained if mixed function oxidases (MFO) and glutathione-s-transferase (GST) activity levels in different tissues are used as surrogates for the (unknown) specific MFO and glutathione resources most relevant for PAH metabolism. In particular, the ratio of MFO specific activities in subcellular preparations of lung tissues from Sprague-Dawley rats compared to preparations of lung tissues from humans is about 0.11/0.0006, while the GST specific activity levels are indistinguishable (Lorenz et al., 1984). Similarly, for mice, the MFO specific activity level is about 0.732 / 0.0006 = 1220 times higher in mice than in humans, while the GST specific activity level is about 727 / 78 = 9.3 times higher. Thus, if the MFO-to-GST ratio is a useful surrogate indicator of species susceptibility based on relative internal doses, then mice should be about two orders of magnitude more susceptible than humans. To a first approximation, in the absence of more specific and detailed information, it might be expected that the rate of formation of carcinogenic PAH metabolites from administered PAH doses via MFO-catalyzed metabolism is also about two orders of magnitude greater in rats and mice than in humans. For chronic, low-level exposures leading to steady-state internal dose concentrations, the internal dose in a rat would be expected to be at least tens, and more probably hundreds, of times greater than the corresponding internal dose in a human, based on the relative specific activities of MFO in the two species.

In summary, the best answer to question Q2 posed above, based on the limited biological evidence available, appears to be the following: The relation between administered and internal doses, under steady-state, low-level exposure conditions, is that average internal dose per unit time is probably proportional to average administered dose per unit time, with the constant of proportionality being about two orders of magnitude greater in mice and in rats than it is in humans.

## DISCUSSION OF SPECIES DIFFERENCES

A default assumption commonly used in regulatory risk assessment is that "mg/surface area/day is an equivalent dose between species" (Collins et al., 1991, p. 175). This default assumption has resulted in adjusting the cancer potency factor for B(a)P estimated from animal data upward by a factor of 12.7 in regulatory risk assessments (based on the relative surface areas of humans

compared to mice) in calculating the corresponding estimated cancer potency for humans (Collins et al., 1991). Modern biologically-based approaches to risk assessment replace this default assumption with the alternative, more specific assumption that average concentration of carcinogenic metabolites in target organs or cell populations per unit time is an equivalent dose between species. Instead of adjusting carcinogenic potency estimated from mouse data upward by a factor of more than 10 to convert to humans, the biological considerations in this section suggest that it might be more appropriate to adjust the mouse potency estimate downward by a factor of 100 or more. This would reduce estimated human carcinogenic potency (based on experimental data from mice) by about three orders of magnitude compared to previous estimates (Collins et al., 1991).

# ESTIMATING THE RISK OF CANCER ASSOCIATED WITH A LOW-LEVEL PAH EXPOSURE

## CONCEPTS FOR CANCER POTENCY CALCULATIONS

As previously stated, the US EPA considers that there is no direct evidence in humans specifically linking B(a)P exposure to increased cancer risks. However, B(a)P is classified by US EPA as a "probable" human carcinogen, primarily because it is known to be a carcinogen in animals at sufficiently high doses. However, a wealth of data suggest that, at the concentrations of interest in the Celotex case, B(a)P has no detectable carcinogenic effect even in animals. Therefore, a final, key source of uncertainty is that *it is not known whether the concentrations of B(a)P found in the soil samples from yards near the industrial property can cause cancer in experimental animals.* If not, then the basis for estimating human cancer risks from these soil concentrations is weakened.

To address questions of carcinogenic potency at low doses, it is usual to assume that at, sufficiently low doses, lifetime probability of tumor is well approximated by a polynomial:

$$\Pr(\text{tumor if dose is } x) = q_0 + q_1 x + q_2 x^2 + \ldots + q_n x^n.$$

For very low doses (as x approaches zero), the behavior of this function is determined by $q_1$. If $q_1$ is positive, then it dominates the dose-response function at low doses: it is just the slope of the dose-response function (measured in units of expected tumors per unit of exposure) at the origin. In this case, $q_1$ is called the potency of the carcinogen. On the other hand, if $q_1 = 0$, then the cancer potency quickly approaches zero at low doses. These are the two qualitatively different behaviors generally considered to be possible for carcinogen dose-response relations in the usual regulatory framework for cancer risk assessment. In the absence of specific evidence to the contrary, the EPA traditionally assumes that carcinogen dose-response functions are "low-dose linear", meaning that $q_1 > 0$.

## ANALYSIS OF LOW-DOSE CANCER POTENCY FOR B(a)P

For B(a)P, there is strong experimental evidence from multiple species and experimental designs that the dose-response function is not low-dose linear. For example, the following table, adopted from the regulatory risk assessment by Collins et al. (1991), shows a clear threshold-like nonlinearity in tumor risk as a function of dose. Between 40 and 50 ppm, the dose-response relation crosses an apparent threshold (or strong upward nonlinearity) above which there is strong carcinogenic potency. Below 40 ppm, carcinogenic potency appears to be weak or non-existent.

TABLE: MICE EXPOSED TO B(a)P BY FEEDING SHOW A NONLINEAR DOSE-RESPONSE RELATION

| Exposure (ppm) | Incidence of gastric tumors |
|---|---|
| 0 | 0 |
| 10 | 0 |
| 20 | 0.043 (= 1 / 23) |
| 30 | 0 |
| 40 | 0.025 (1 / 40) |
| 45 | 0.10 (4 / 40) |
| 50 | 0.71 (24 / 34) |
| 100 | 0.83 (19 / 23) |

Other data sets, for mice exposed to B(a)P by subcutaneous injection (Bryan and Shimkin, 1943) or skin painting (Wynder and Hoffman, 1959), for Syrian Golden hamsters exposed by inhalation (Thyssen et al., 1981), and so forth show a similar pattern of a relatively abrupt, well-localized transition from no significant tumor risk to very high tumor risk as concentration increases by less than one order of magnitude.

Such data provide strong evidence of a dose-response relation that is nonlinear at low doses (e.g., at doses for which the response probability is less than 0.01.) Indeed, in order to fit low-dose linear risk models, regulators have in the past had to discard some of the high-dose data (Collins et al., 1991). A

formal statistical test of the null hypothesis of low-dose linearity would reject it overwhelmingly in favor of the alternative hypothesis of low-dose nonlinearity.

The implications of low-dose nonlinearity for risk assessment can be dramatic. For example, the dose-response data in the preceding table suggest that there is no detectable excess tumor risk at ingested B(a)P concentrations below about 30 ppm in mice. If humans are less susceptible than mice, as suggested by comparing enzyme activity levels across species, then no excess cancer risk would be expected at concentrations of 30 ppm or less in humans, based on the mouse data. Only by ignoring some of the data points and applying default assumptions and risk models that have not been customized to reflect B(a)P-specific data is it possible to reach the opposite conclusion, that there are excess cancer risks at concentrations below 30 ppm. In particular, regulatory risk assessments of B(a)P have assumed low-dose linearity, rather than treating it as a hypothesis to be tested based on experimental data. If the hypothesis of low-dose nonlinearity is accepted, as the data seem to require, then both the expected risk and upper confidence bands on expected risk approach zero at low doses (e.g., 30 ppm and below).

## DISCUSSION OF REGULATORY DEFAULT ASSUMPTIONS FOR B(a)P

The default assumptions used in regulatory risk assessments of B(a)P guarantee a positive answer to the question of whether B(a)P at low doses creates excess risk, by assuming low-dose linearity. But low-dose linearity is contradicted by all of the experimental data, suggesting that the default assumption is not appropriate for B(a)P. Some of the other default assumptions should also be revised. For example, the estimated cancer potency in mice based on the empirically observed proportions of mice developing tumors was arbitrarily multiplied by a factor of 40 to reflect an assumption "that cancer incidence increases as the third power of age" (Collins et al., 1991). This is clearly inappropriate when mice at the higher dose levels already have tumor incidence rates in excess of 70%. Assuming that 40 times as many mice would have developed tumors had the experiment been continued longer is incoherent. Thus, it appears that a thorough review of default assumptions is in order for B(a)P, and that several of the assumptions must be refined or replaced

in order to achieve a more realistic assessment of B(a)P-induced cancer risks at concentrations below about 30 ppm in food.

At the industrial property, humans are exposed to orders of magnitude lower doses than those in the mouse experiment. Therefore, any plausible low dose nonlinear dose-response function would predict no significant excess risks to humans based on their exposures to PAHs in soil. The best estimate of absolute risk due to soil PAHs is that it is indistinguishable from zero.

## DISCUSSION OF PAHs OTHER THAN B(a)P

Most of this appendix has concentrated on B(a)P as a surrogate for other PAHs. It is worth considering whether the complex mixture of soil PAHs might present a greater cancer risk than would B(a)P alone. However, there is evidence (e.g., Cherng et al., 1996; Springer et al., 1989) that antagonism among the PAHs is more likely than synergy. Also, at the low concentrations involved, any such interactions are likely to be weak. Therefore, the two main conclusions are not changed by considering that there are multiple PAHs. It is still the case that (i) Previous risk assessments for B(a)P have used default assumptions that are not appropriate for B(a)P; and (ii) The empirical dose-response data for B(a)P suggests that soil PAHs contribute negligibly to cancer risks in the vicinity of the industrial property.

## REFERENCES

Cavalieri, E.L., and E.G. Rogan, 1992. The approach to understanding aromatic hydrocarbon carcinogenesis: The central role of radical cations in metabolic activation. *Pharmac. Therap.* 55, 183-199.

Collins, J.F., J.P. Brown, S.V. Dawson, and M.A. Marty, 1991. Risk assessment for benzo(a)pyrene. *Regulatory Toxicology and Pharmacology.* 13, 170-184.

Cherng, S-H, S.T. Lin, and H. Lee, 1996. Modulatory effects of polycyclic aromatic hydrocarbons on the mutagenicity of 1-nitropyrene: A structure-activity relationship study. *Mutation Research,* 367, 177-185.

Cox, L.A., Jr., "Simple relations between administered and internal doses in compartmental flow models," *Risk Analysis,* **15**, 2, 197-204, 1995.

Doll, R., and R. Peto, 1981. *The Causes of Cancer.* Oxford University Press.

Grinberg-Funes, R.A., et al., 1994. Polycyclic aromatic hydrocarbon - DNA adducts for smokers and their relationship to micronutrient leveels and the glutathione-S-transferase M1 genotype. *Carcinogenesis,* 15, 11, 2449 - 2454.

Kirby, G.M., et al. 1995. "Influences of chronic cholangiohepatitis and cholestasis on hepatic metabolism of benzo(a)pyrene in white suckers from industrially polluted areas of lake Ontario," *Carcinogenesis,* 16, 12, 2923-2929.

Lorenz, J., H.R. Glatt, et al., 1984. Drug metabolism in man and its relationship to that in three rodent species. *Biochem. Med.,* 32, 43-56.

Springer, D.I., et al., 1989. Influences of complex organic mixtures on tumor-promoting activity, DNA binding and adducts of benzo(a)pyrene. *Carcinogenesis,* 10, 1, 131-137.

# Appendix I

# Copies of Peer-Reviewed Publications

25 October 1996

Alceon Corporation
PO Box 382669
Harvard Square Station
Cambridge, MA 02238-2669

617-864-4300

# Appendix I

## Copies of Peer Reviewed Publications

Brainard & Burmaster, 1992
Brainard, J. and D.E. Burmaster, 1992, Bivariate Distributions for Height and Weight of Men and Women in the United States, Risk Analysis, 1992, Volume 12, Number 2, pp 267 - 275

Burmaster, Lloyd & Crouch, 1994
Burmaster, D.E., K.J. Lloyd, and E.A.C. Crouch, 1994, LogNormal Distributions of Body Weight as a Function of Age for Female and Male Children in the United States, In Revision, Risk Analysis, in revision

Burmaster, Lloyd & Thompson, 1995
Burmaster, D.E., K.J. Lloyd, and K.M. Thompson, 1995, The Need for New Methods to BackCalculate Soil CleanUp Targets in Interval and Probabilistic Risk Assessments, Human and Ecological Risk Assessment, Volume 1, Number 1, pp 89 - 100

Burmaster & Thompson, 1995
Burmaster, D.E. and K.M. Thompson, 1995, BackCalculating CleanUp Targets in Probabilistic Risk Assessments When the Acceptability of Cancer Risk is Defined Under Different Risk Management Policies, Human and Ecological Risk Assessment, Volume 1, Number 1, pp 101 - 120

Burmaster & Anderson, 1994
Burmaster, D.E. and P.D. Anderson, 1994, Principles of Good Practice for the Use of Monte Carlo Techniques in Human Health and Ecological Risk Assessments, Risk Analysis, Volume 14, Number 4, pp 477 - 481

Burmaster & von Stackelberg, 1991
Burmaster, D.E. and K. von Stackelberg, 1991, Using Monte Carlo Simulations in Public Health Risk Assessment: Estimating and Presenting Full Distributions of Risk, Journal of Exposure Analysis and Environmental Epidemiology, Volume 1, Number 4, pp 491 - 512

Thompson et al, 1992
Thompson, K.M., D.E. Burmaster, and E.A.C. Crouch, 1992, Monte Carlo Techniques for Quantitative Uncertainty Analysis in Public Health Risk Assessments, Risk Analysis, Volume 12, Number 1, pp 53 - 63

Thompson & Burmaster, 1991
Thompson, K.M. and D.E. Burmaster, 1991, Parametric Distributions for Soil Ingestion by Children, Risk Analysis, Volume 11, Number 2, pp 339 - 342

Burmaster & Bloomfield, 1996
Burmaster, D.E. and L.R. Bloomfield, 1996, Mathematical Properties of the Risk Equation When Variability is Present, Human and Ecological Risk Assessment, Volume 2, Number 2, pp 348 - 355

Burmaster, 1996
Burmaster, D.E., 1996, Using Beta Distributions Efficiently in Probabilistic Exposure Assessments, Risk Analysis, in review

Risk Analysis, Vol. 12, No. 2, 1992

# Bivariate Distributions for Height and Weight of Men and Women in the United States

Jennifer Brainard[1] and David E. Burmaster[1]

Received May 21, 1991; revised August 16, 1991

For the U.S. population, we fit bivariate distributions to estimated numbers of men and women aged 18–74 years in cells representing 1 in. intervals in height and 10 lb intervals in weight. For each sex separately, the marginal histogram of height is well fit by a normal distribution. For men and women, respectively, the marginal histogram of weight is well fit and satisfactorily fit by a lognormal distribution. For men, the bivariate histogram is satisfactorily fit by a normal distribution between the height and the natural logarithm of weight. For women, the bivariate histogram is satisfactorily fit by two superposed normal distributions between the height and the natural logarithm of weight. The resulting distributions are suitable for use in public health risk assessments.

KEY WORDS: Height; body weight; univariate; bivariate; distribution; simulation.

## 1. INTRODUCTION

For many years, people analyzing public health risks at or near hazardous waste sites have assumed that all adults weigh 70 kilograms (kg), although some analysts have assumed different weights for men and women. Point estimates now appear routinely as standard assumptions in guidance manuals published by the U.S. Environmental Protection Agency (EPA) for the "Superfund" and related programs (e.g., Ref. 1). More recently, the EPA has published a simple table of arithmetic means and standard deviations for body weights of men, women, and men and women together in different age groups (Ref. 2, p. 5–5). This most recent Agency approach stops considerably short of the continuous curves of mean body weights (with error bars) reported in the "Report of the Task Group on Reference Man."[3]

In this manuscript, we examine data on the height and weight of adults published by the U.S. Public Health Service and fit bivariate distributions to the tabulated values for men and women separately. Based on the second National Health and Nutrition Examination Survey (NHANES II), conducted from February 1976 through February 1980, the U.S. Public Health Service has published extensive tables of heights and weights of the U.S. civilian noninstitutionalized population from six months to 74 years of age.[4] In the field survey, trained examination teams tabulated the height and weight of 5916 men and 6588 women in the age range 18–74 years. After statistically adjusting the raw data to reflect the whole U.S. population aged 18–74 years with regard to age structure, sex, and race, the U.S. Public Health Service published the results shown in Table I (for an estimated 67,552 thousand men) and Table II (for 74,167 thousand women). (In the original publication,[4] Tables 27 and 28 suffer from minor discrepancies in the marginal counts, corrected here by resumming the rows and columns.)

Tables I and II, respectively, report the estimated number of men and women in the U.S. in the age range of 18–74 years grouped in cells representing 1 in. intervals in height and 10 lb intervals in weight. As expected, the (adjusted) data show (i) that men are taller and weigh more, on average, than women, and (ii) that taller people of either sex, on average, weight more than shorter peo-

[1] Alcoon Corporation, P. O. Box 2669, Harvard Square Station, Cambridge, Massachusetts 02238-2669.



0272-4332/92/0600-0267$06.50/1 

Table II. Number of Women 18–74 Years of Age, by Weight and Height, Unites States, 1976–1980 (Number of Persons in Thousands)[a,d]

| Height[b] (in) | Weight[c] <90 (lb) | 90–99 (lb) | 100–109 (lb) | 110–119 (lb) | 120–129 (lb) | 130–139 (lb) | 140–149 (lb) | 150–159 (lb) | 160–169 (lb) | 170–179 (lb) | 180–189 (lb) | 190–199 (lb) | 200–209 (lb) | 210–219 (lb) | ≥220 (lb) | True total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| <55 | | 7 | 8 | 11 | 25 | | 7 | 7 | 15 | | | | | | | 80 |
| 55 | | | 13 | | 4 | 26 | 12 | 31 | 13 | 8 | | | | | | 107 |
| 56 | 31 | 57 | 12 | 41 | 55 | 25 | 44 | 25 | | | | 6 | | | | 296 |
| 57 | 44 | 91 | 107 | 90 | 55 | 115 | 76 | 26 | 24 | | 9 | 18 | | 4 | 36 | 695 |
| 58 | 93 | 164 | 132 | 338 | 317 | 147 | 78 | 120 | 35 | 68 | 27 | 14 | 34 | 3 | 42 | 1612 |
| 59 | 50 | 196 | 262 | 552 | 342 | 365 | 297 | 201 | 123 | 116 | 69 | 46 | 30 | | 31 | 2680 |
| 60 | 86 | 267 | 538 | 621 | 722 | 775 | 451 | 334 | 261 | 239 | 128 | 99 | 54 | 30 | 40 | 4645 |
| 61 | 12 | 368 | 754 | 1286 | 1355 | 1089 | 877 | 807 | 439 | 308 | 269 | 240 | 123 | 110 | 164 | 8201 |
| 62 | 14 | 258 | 938 | 1660 | 1899 | 1306 | 1117 | 728 | 583 | 448 | 305 | 227 | 130 | 117 | 218 | 9948 |
| 63 | 32 | 165 | 843 | 1729 | 1776 | 1600 | 1565 | 1006 | 817 | 655 | 477 | 357 | 277 | 151 | 283 | 11,733 |
| 64 | | 30 | 531 | 1168 | 1653 | 1936 | 1475 | 950 | 741 | 513 | 404 | 274 | 117 | 198 | 280 | 10,270 |
| 65 | | 64 | 283 | 873 | 1582 | 2162 | 1183 | 1201 | 693 | 396 | 455 | 269 | 156 | 109 | 516 | 9942 |
| 66 | | 10 | 76 | 705 | 804 | 1365 | 902 | 696 | 509 | 255 | 193 | 213 | 116 | 84 | 253 | 6181 |
| 67 | | | 32 | 188 | 514 | 740 | 605 | 336 | 338 | 381 | 275 | 155 | 106 | 67 | 253 | 3990 |
| 68 | | | 10 | 85 | 213 | 488 | 369 | 336 | 193 | 41 | 99 | 95 | 82 | 14 | 106 | 2131 |
| 69 | | | 33 | | 98 | 135 | 266 | 125 | 214 | 119 | 43 | 28 | | | 93 | 1154 |
| 70 | | | | | 6 | 38 | 56 | 52 | 19 | 46 | | 25 | 3 | | | 245 |
| ≥71 | | | | 16 | | 16 | 55 | 42 | 30 | 28 | | 15 | 4 | | 51 | 257 |
| True total | 362 | 1677 | 4572 | 9363 | 11,420 | 12,328 | 9435 | 7023 | 5047 | 3621 | 2753 | 2081 | 1232 | 887 | 2366 | 74,167 |

[a] *Source*: Ref. 4, Table 28.
[b] Height without shoes.
[c] Weight with clothes, estimated as ranging from 0.20–0.62 lb.
[d] Numbers in cells scaled up to reflect size of population; only 10,339 women actually examined.

$z$-score. We used ordinary least squares to fit the best straight lines through the appropriate variables.

### 3.2. Results for the Marginal Distributions

For men, Figs. 1 and 2, respectively, show the marginal cumulative values and associated $z$-scores for $Ht$ and $\ln Wt$. The straight lines for both $Ht$ and $\ln Wt$, fit to the points by ordinary linear regression, have $R^2$ values of 0.999. From the intercepts and slopes of the regression lines,[9] we estimate values (i) for $\mu_{Ht}$ and $\sigma_{Ht}$ and (ii) for $\mu_{\ln Wt}$ and $\sigma_{\ln Wt}$ as shown in Table III. The excellent visual fits and the high $R^2$ values for the best-fit line for $Ht$ and $\ln Wt$ support the inference that the marginal distributions for $Ht$ and $\ln Wt$ for men are both Gaussian in form.

For women, Figs. 3 and 4, respectively, show the marginal cumulative values and associated $z$-scores for $Ht$ and $\ln Wt$. These straight lines, also fit to the points by ordinary linear regression, have $R^2$ values of 0.999 and 0.985, respectively, for $Ht$ and $\ln Wt$. From the intercepts and slopes of the regression lines, we estimate values for the four parameters as shown in Table III. The close visual fit and the high $R^2$ value for the best-fit line for $Ht$ for women support the inference that the marginal distribution is Gaussian in form. The corresponding inference for $\ln Wt$ for women is weaker but adequate.

## 4. CHARACTERIZATION OF BIVARIATE DISTRIBUTIONS

### 4.1. Methodology for the Bivariate Distributions

As a first step in fitting a bivariate distribution to $Ht$ and $\ln Wt$ for each sex, we estimated the Pearson (linear) correlation coefficient (denoted $\rho$) and the Spearman (rank) correlation coefficient (denoted $\rho_{rank}$) between $Ht$ and $\ln Wt$ using the observed binned data for men and women.

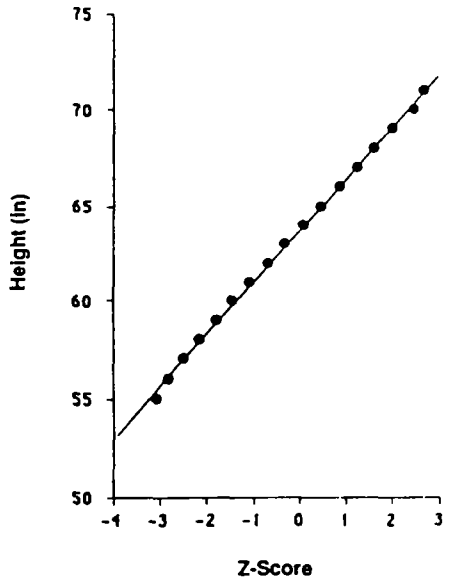



Fig. 3. Women: height vs. z-score.

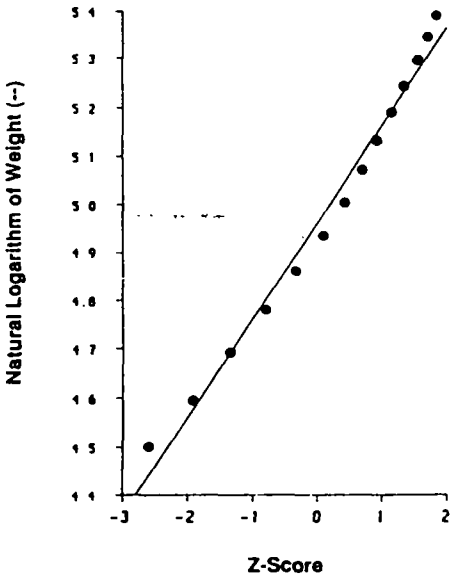



Fig. 4. Women: natural log of weight vs. z-score.

Table III. Estimated Parameters for Univariate and Bivariate Normal Distributions

| | Men | | | | Women | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Source: | Variable | Estimated μ | Estimated σ | Estimated ρ | Variable | Estimated μ | Estimated σ | Estimated ρ | Fraction from each distribution |
| Marginals analysis | *Ht* | 69.12 | 2.85 | | *Ht* | 63.68 | 2.68 | | |
| | ln*Wt* | 5.13 | 0.17 | | ln*Wt* | 4.96 | 0.20 | | |
| Pearson correlation | | | | 0.38 | | | | 0.22 | |
| Spearman correlation | | | | 0.37 | | | | 0.22 | |
| Minimization of $\chi^2$ statistic assuming one distribution each for men and women | *Ht* | 69.18 | 2.87 | 0.42 | *Ht* | 63.81 | 2.68 | 0.24 | |
| | ln*Wt* | 5.14 | 0.17 | | ln*Wt* | 4.95 | 0.21 | | |
| Minimization of $\chi^2$ statistic assuming two distributions for women | | | | | $Ht^a_{(1)}$ | 63.11 | 2.76 | | |
| | | | | | | | 0.41 | 0.46 | |
| | | | | | $\ln Wt^a_{(1)}$ | 5.06 | 0.24 | | |
| | | | | | $Ht^b_{(2)}$ | 64.36 | 2.49 | | |
| | | | | | | | | 0.44 | 0.54 |
| | | | | | $\ln Wt^b_{(2)}$ | 4.86 | 0.14 | | |

[a] These parameters are for the first of two distributions for women.
[b] These parameters are for the second of two distributions for women.

eters are similar to those estimated earlier by marginal analysis and by Pearson and Spearman correlations for the noncensored data. From the bivariate analysis, we estimate the arithmetic average for height and weight as 69.2 in. and 173.2 lb, respectively.

For women, a similar calculation with a single bivariate normal distribution gave poor results in terms of the total $\chi^2$ statistic and in terms of patterns in the optimized residuals. Table III shows the optimized parameters from this calculation, but we consider them less useful in practice. The five optimized parameters are also similar to those estimated earlier by marginal analy-

Table V. Optimized Residuals for Each Bin for Women Assuming Two Distributions*

| Height (in) | LnWeight 3.81 (lb) | 4.01 (lb) | 4.17 (lb) | 4.32 (lb) | 4.44 (lb) | 4.55 (lb) | 4.65 (lb) | 4.74 (lb) | 4.83 (lb) | 4.91 (lb) | 4.98 (lb) | 5.04 (lb) | 5.11 (lb) | 5.16 (lb) | 5.22 (lb) | 5.27 (lb) | 5.32 (lb) | 5.37 (lb) | 5.42 (lb) | 5.46 (lb) | 5.50 (lb) | 5.54 (lb) | 5.58 (lb) | 5.62 (lb) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 51.5 | | | | | | | | | | | | | | | | | | | | | | | | |
| 52.5 | | | | | | | | | | | | | | | | | | | | | | | | |
| 53.5 | | | | | | | | | | | | | | | | | | | | | | | | |
| 54.5 | | | | | −7 | −1 | −2 | 2 | 18 | | | | | | | | | | | | | | | |
| 55.5 | | | | | −11 | −15 | −6 | −20 | −13 | 13 | 2 | 24 | 9 | 5 | −1 | −1 | 0 | 0 | | | | | | |
| 56.5 | | | | | 10 | 23 | −37 | −12 | 8 | −13 | 16 | 5 | −13 | −8 | −5 | 3 | −2 | −1 | | | | | | |
| 57.5 | | | | | 8 | 21 | −5 | −35 | −58 | 22 | 4 | −26 | −12 | −24 | −6 | 9 | −5 | 1 | | | | | | |
| 58.5 | | | | | 41 | 40 | −90 | 73 | 73 | −53 | −78 | 2 | −50 | 9 | −12 | −10 | 19 | −6 | | | | | | |
| 59.5 | | | | | −15 | 7 | −124 | 52 | −133 | −21 | −3 | −29 | −48 | −6 | −15 | −9 | −5 | −22 | 2 | | | | | |
| 60.5 | | | | | 16 | 21 | −42 | −208 | −104 | 103 | −60 | −55 | −35 | 20 | −28 | −8 | −16 | −15 | −25 | | | | | |
| 61.5 | | | | | −51 | 97 | 11 | 97 | 86 | 39 | 95 | 225 | −4 | −27 | 22 | 64 | 2 | 30 | 39 | | | | | |
| 62.5 | | | | | −35 | 7 | 138 | 209 | 207 | −153 | 40 | −48 | 3 | 4 | −31 | −21 | −46 | −4 | 15 | | | | | |
| 63.5 | | | | | 1 | −29 | 122 | 239 | −150 | −170 | 240 | 78 | 146 | 145 | 83 | 58 | 57 | −6 | 0 | | | | | |
| 64.5 | | | | | −17 | −94 | −8 | −108 | −193 | 93 | 40 | −44 | 49 | 0 | 7 | −35 | −118 | 24 | −56 | | | | | |
| 65.5 | | | | | −8 | −2 | −51 | −33 | 106 | 540 | −160 | 257 | 55 | −60 | 106 | −6 | −59 | −55 | 173 | | | | | |
| 66.5 | | | | | −3 | −19 | −94 | 174 | −175 | 171 | −167 | −85 | −14 | −105 | −75 | 2 | −53 | −49 | −46 | | | | | |
| 67.5 | | | | | −1 | −11 | −40 | −68 | −22 | 11 | −109 | −217 | −39 | 129 | 94 | 14 | −8 | −25 | 30 | | | | | |
| 68.5 | | | | | 0 | −3 | −15 | −17 | −29 | 120 | −28 | 5 | −41 | −115 | −9 | 13 | 16 | −40 | −36 | | | | | |
| 69.5 | | | | | | −1 | 26 | −33 | 9 | −18 | 83 | −41 | 90 | 35 | −14 | −14 | −33 | −28 | 15 | | | | | |
| 70.5 | | | | | | 0 | −2 | −9 | −21 | −14 | −14 | −17 | −36 | 7 | −26 | 6 | −12 | −12 | −36 | | | | | |
| 71.5 | | | | | | | | 14 | −8 | −3 | 26 | 10 | 1 | 6 | −15 | 4 | −4 | −7 | 29 | | | | | |
| 72.5 | | | | | | | | | | | | | | | | | | | | | | | | |
| 73.5 | | | | | | | | | | | | | | | | | | | | | | | | |
| 74.5 | | | | | | | | | | | | | | | | | | | | | | | | |
| 75.5 | | | | | | | | | | | | | | | | | | | | | | | | |

* Optimized residual refers to the difference between the predicted and observed values in a bin obtained when the sum of the chi-squares is minimized.

We calculated the observed marginal $\chi^2$ values by summing the observed and expected binned data (obtained from minimizing the total $\chi^2$ statistic for the cumulative distributions) across *Ht* or ln*Wt* and by calculating the marginal optimized residuals, the marginal $\chi^2$ values, and the sum of the marginal $\chi^2$ values. For example, we obtained the $\chi^2$ value for height between 65 and 66 in. by summing the observed and expected binned data for this height interval across all weight categories. The resulting sum of the observed binned data minus that for the expected equals the optimized residual for the 65–66 in. height interval. We then estimated the total marginal $\chi^2$ as, for example,

$$\sum \frac{(O_i - E_i)^2}{E_i}$$

over the height intervals.

The final $\chi^2$ statistic for each marginal distribution may be compared with the $\chi^2_{0.05}$ value with degrees of freedom equal to the number of *Ht* or ln*Wt* categories minus one.[15] For both men and women, the observed $\chi^2$ values, the degrees of freedom, the $\chi^2_{0.05}$ values, and the *p*-values for the observed $\chi^2$ values for the marginal analyses are shown in Table VI. Because all of the observed $\chi^2$ values for men exceed their respective $\chi^2_{0.05}$ values (i.e., *p*-values < 0.05), we reject the null hypothesis and conclude that the observed and predicted distributions do not come from the same population. Assuming a single distribution for women, we reject the null hypothesis for *Ht* and ln*Wt* since the observed $\chi^2$ values also exceed their respective $\chi^2_{0.05}$ values (i.e., *p*-values < 0.05). Assuming two distributions for women, the observed $\chi^2$ value for *Ht* exceeds its $\chi^2_{0.05}$ value, while the observed $\chi^2$ value for ln*Wt* is less than its $\chi^2_{0.05}$ value (i.e., *p*-value > 0.05). Consequently, we reject

ifies the simulation for men, the simulation for women requires the use of a Bernoulli trial to select between two distributions, one for each subpopulation.

## 7. SUMMARY AND DISCUSSION

Bivariate data for the height and weight of men and women between the ages of 18 and 74 years are well fit by normal distributions between the height and the logarithm of weight. For men, a single bivariate normal distribution fits the data well, and for women, a pair of superposed bivariate normal distributions fits the data well. The final distributions of height and weight for men and women are suitable and practical for use in public health risk assessments using Monte Carlo simulation to estimate full distributions for exposure and risk.

## ACKNOWLEDGMENTS

Many people have contributed to this reasearch. We give special thanks to William Rand of Tufts University who advised and encouraged us many times throughout a year of effort. We also thank, in alphabetical order, Paul Abbott of Wolfram Research Inc., Brian Junker of Carnegie-Mellon University, Nancy I. Maxwell of Alceon Corporation, Michael Rowland of National Center for Health Statistics, Bo E.H. Saxberg of the Massachusetts Institute of Technology, Lawrence Seiford of the University of Massachusetts at Amherst, and Cameron Smith of Variable Symbols, Inc., for their helpful suggestions.

Alceon Corporation funded this research.

## REFERENCES

1. U.S. Environmental Protection Agency, *Risk Assessment Guidance for Superfund, Vol. I, Human Health Evaluation Manual (Part A)* (EPA/540/1-89/002, Office of Emergency and Remedial Response, Washington, D.C., December 1989).
2. U.S. Environmental Protection Agency, *Exposure Factors Handbook* (Final Report, EPA/600/8-89/043, Office of Health and Environmental Assessment, Washington, D.C. March 1989).
3. W. S. Snyder, M. J. Cook, E. S. Nasset, L. R. Karhausen, G. P. Howells, and I. H. Tipton, "Report of the Task Group on Reference Man," *International Commission on Radiological Protection Number 23* (Pergamon Press, Oxford, U.K., 1975).
4. National Health Survey, "Anthropometric, Reference Data and Prevalence of Overweight, United States, 1976–80" (Data from the National Health Survey Series, Number 238, DHHS Publication PHS 87-1688, U.S. Department of Health and Human Services, Hyattsville, Maryland, October 1987).
5. J. W. Tukey, *Exploratory Data Analysis* (Addison-Wesley, Reading, Massachusettes, 1977).
6. J. M. Chambers, *Computational Methods for Data Analysis* (John Wiley & Sons, New York, 1977).
7. $D^2$ Software, Inc., *MacSpin Graphical Data Analysis Software* (release notes for Version 3.0, Austin, Texas, 1990).
8. Systat, Inc., *Users Manuals for SYGRAPH Graphics and SYSTAT Statistics* (Evanston, Illinois, 1990).
9. C. C. Travis and M. Land, "Estimating the Mean of Data Sets with Nondetectable Values," *Environmental Science and Technology* 24, 961–962 (1990).
10. J. E. Freund, *Mathematical Statistics*, 2nd ed. (Prentice-Hall, Englewood Cliffs, New Jersey, 1971).
11. L. S. Lasdon, A. D. Warren, A. Jain, and M. Ratner, "Design and Testing of a Generalized Reduced Gradient Code for Nonlinear Programming," *Transactions on Mathematical Software*, 4, 34–50 (1978).
12. W. Wasil, B. Golden, and L. Liu, "State-of-the Art in Nonlinear Optimization Software for the Microcomputer," *Computers Operations Research* 16, 497–512 (1989).
13. Microsoft Corporation, *Microsoft Excel User's Guide* (Redmond, Washington, 1991).
14. W. M. Rand, (personal communication, January 1991).
15. D. L. Harnett, *Introduction to Statistical Methods*, 2nd ed. (Addison-Wesley, Reading, Massachusettes, 1975).
16. T. P. Hutchinson and C. D. Lai, *Continuous Bivariate Distributions, Emphasizing Applications* (Rumsby Scientific Publishing, Adelaide, South Australia, 1990).
17. Market Engineering Corporation, *User's Guide for Version 2.0* (Denver, Colorado, 1991).

~~Submitted to Risk Analysis~~
In Revision

# Lognormal Distributions of Body Weight as a Function of Age for Female and Male Children in the United States

David E. Burmaster, Ph.D.
Kylie J. Lloyd, M.A., M.P.H.
Alceon Corporation
PO Box 2669
Harvard Square Station
Cambridge, MA 02238-2669
617-864-4300
deb@Alexandria.LCS.MIT.EDU

with an Appendix by
Edmund A.C. Crouch, Ph.D.
Cambridge Environmental, Inc.
58 Charles Street
Cambridge, MA 02141
617-225-0810

## Abstract

In both deterministic and probabilistic human health risk assessments, body weight plays a crucial role in estimating exposure doses and the subsequent health risks. Based on results published from the NHANES II Survey (National Health and Nutrition Examination Survey, 1989), we use exploratory data analysis, probability plots, and regressions to fit normal and lognormal distributions to percentiles of body weight for female and male children as a function of age from 6 months to 20 years. Lognormal distributions give consistently strong fits to the NHANES II data across all age groups for each gender, a result consistent with previously published results for adult women and men. We also demonstrate the practical use of these results in risk assessments.

## Introduction

In 1983, the National Academy of Science published a method often used to estimate the health risks associated with exposure to hazardous chemicals in the environment (NAS, 1983). In a deterministic risk assessment, an analyst combines point values representing the intensity, frequency, and duration of exposure with point values for toxicity to estimate a health risk. Each of these point values may be an average, conservative, upper-bound, or worst-case value. We have found that risk assessors often estimate health risks for exposure scenarios that will rarely if ever occur because

the combination of point values often falls far above the 95th percentile of the full range (Burmaster & Harris, 1993).

Using probabilistic techniques, a risk assessor can estimate full distributions of exposure and risk. In a probabilistic risk assessment, the analyst specifies a probability density function - PDF (or a cumulative distribution function - CDF) for each input variable to estimate full distributions for exposure and risk.

To conduct a probabilistic risk assessment that includes children in the exposed population, a risk assessor needs parametric or nonparametric distributions for the body weights of children as a function of age. The US EPA has published averages for children's body weights grouped in three-year intervals; birth to 3 years, 3 years to 6 years, 6 years to 9 years and so on to 18 years (US EPA, 1989, EFH, Table 5-3), but the Agency's manual does not give full distributions for the body weights as a function of age. As children's body weights change with age, many analysts working on deterministic risk assessments return to the US EPA's source, the NHANES II Survey completed by the National Center for Health Statistics (NCHS, 1987), to use the data in single year increments. By extension, we return to the same source to fit probability distributions for use in probabilistic risk assessments.

**NHANES II Survey Design**

The NHANES II Survey collected information on the nutritional status and related factors to determine the prevalence of overweight people in the United States (US) non-institutionalized population. Conducted by the National Center for Health Statistics (NCHS) from February 1976 through February 1980, the target population was civilians in 50 states from 6 months through 74 years of age.

The entire NHANES II sample included 27,801 persons, 91 percent of whom were interviewed. Of these, 20,322 were interviewed and examined, for a response rate of

73.1 percent (NCHS, 1987). The body weights of 4,079 females and of 4,379 males less than 20 years of age were collected and reported after the data were statistically adjusted for non-response and probability of selection and then post-stratified by age, sex, and race to reflect the whole US population (Exhibit 1; NCHS, 1987).

As described in Appendix 1 of the NHANES II Survey, the survey used a stratified, multistage design that selected samples at each stage with a known probability of sampling females and males. In hierarchical order, the stages of selection were: primary sampling units (PSUs), which are counties or small groups of contiguous counties; census enumeration districts; segments (clusters of households); households; and finally sample persons. The list consisted of all housing units located in the 1970 Census of the Population (NCHS, 1987). Younger and older age groups were over-sampled and approximately one person per sample household was selected.

The NCHS derived national estimates through a multistage estimation procedure with three main steps: (i) inflation by the reciprocal of the probability of selection, (ii) adjustment for non-response, and (iii) post-stratification by age, sex, and race. (NCHS, 1987, Appendix I - Statistical Notes). The probability of selection is the product of the probabilities of selection from each stage of selection in the design - PSU, segment, household, and sample person to reduce inflation by the reciprocal of the probability of selection (NCHS, 1987, Appendix I - Statistical Notes). To adjust for non-responses, the estimates were inflated by a multiplication factor that increases the estimates based on examined persons to the value that would have been achieved if all sample persons had been examined (NCHS, 1987, Appendix I - Statistical Notes). To post-stratify by age, sex, and race, estimates of the number of examined persons were adjusted by ratio within each of the 75 age-sex-race cells to independent estimates, provided by the US Bureau of the Census, of the population for 1 March 1978, the approximate midpoint of the survey. The ratio adjustment used a multiplication factor in

which the numerator was the US population and the denominator was the sum of the weights adjusted for non-response for examined persons bringing the population estimates into close agreement with the US Bureau of the Census estimates of the civilian non-institutionalized population (NCHS, 1987, Appendix I - Statistical Notes).

**Exploratory Data Analysis**

The NCHS reported the NHANES II results as percentiles of body weight (pounds) for each age group and each gender. (Each age group begins on the birthday of the child and continues for 364 days). In Exhibits 1A and 1B, the percentiles of body weight (converted to kilograms, kg) are almost the same for females and males from 6 months to approximately 10 years of age. As expected, the differences between females and males widen at puberty. For each age group and each gender, Exhibits 1A and 1B show the number of children examined, the mean body weight for the single year increment, and body weights for 9 percentiles: 5th; 10th; 15th; 25th; 50th; 75th; 85th; 90th; and the 95th.

Exhibit 2 graphs female's and male's body weights as a function of age (values are plotted at the mid point of the age group). We note that the 95th percentiles of body weight for the older groups of children are further from the median than the 5th percentiles, thereby indicating positively skewed distributions (Chambers et. al., 1983; Cleveland, 1985; and Tukey, 1977). We also note that all age groups (except the first) have a mean value larger than the median, another indication of positively skewed distributions. The panels in Exhibit 2 also reveal the relatively large sampling errors and statistical fluctuations in the lower and higher percentiles for ages greater than 10 years.

We used Microsoft Excel™ to graph the empirical CDFs for female's and male's body weight on both linear and logarithmic scales. (Throughout this analysis, we use natural logarithms). As an example for one age group, Exhibit 3 shows the percentile body

weights for females and males from 12 to 13 years of age. (As we demonstrate later, females and males at this age have the greatest variance in their weights, i.e., these age groups are the least homogeneous for each gender.) In general, if percentile values come from a normal distribution, then the empirical CDF on a linear scale will have a symmetric sigmoid shape. Similarly, if percentile values come from a lognormal distribution, then the empirical CDF on a logarithmic scale will have a symmetric sigmoid shape. It is usually impossible, however, to determine how well a set of data or percentiles fits a distribution by looking at an empirical CDF. The linear interpolations of the CDFs on a linear scale in the top panel of Exhibit 3 do not appear to be significantly more or less symmetric than those on a logarithmic scale in the bottom panel. By graphical inspection alone, we could not determine whether the normal or lognormal distributions fit the NHANES II percentiles better.

**The Probability Models**

To model the percentiles of body weight, we investigated several symmetric and asymmetric distributions and then focused on the normal distribution and the lognormal distribution. For the normal model, we used this form:

$$\mathrm{BWt} \sim \mathrm{N}(\mu_1, \sigma_1) \qquad \text{Eq 1}$$

where the parameters $\mu_1$ and $\sigma_1$ correspond, respectively, to the mean and the standard deviation of the distribution. For the lognormal model, we used this form:

$$\ln \mathrm{BWt} \sim \mathrm{N}(\mu_2, \sigma_2) \qquad \text{Eq 2}$$

where $\mu_2$ and $\sigma_2$ have the corresponding meaning for the natural logarithm of the distribution. Many texts present material on these distributions (e.g., Evans, 1993).

## Probability Plots and Regression Analyses

For each age group and gender, we plotted the percentile body weights on both normal and lognormal probability plots to compare quantitatively how well the normal and lognormal models fit the data. By design, a normal probability plot has the property that when percentiles or data from a normal distribution are graphed, the points will fall along a straight line (with the intercept equal to $\hat{\mu}_1$ and the slope equal to $\hat{\sigma}_1$) (D'Agostino & Stephens, 1986).

Normal Model: $$BWt = \mu_1 + \sigma_1 \bullet z \qquad \text{Eq 3}$$

Similarly, a lognormal probability plot has the corresponding property that when percentiles or data from a lognormal distribution are graphed, the points will fall along a straight line (with the intercept equal to $\hat{\mu}_2$ and the slope equal to $\hat{\sigma}_2$).

Lognormal Model: $$\ln BWt = \mu_2 + \sigma_2 \bullet z \qquad \text{Eq 4}$$

In each case, the abscissa plots the z-score, equivalent to the variate of a standardized (unit) normal distribution (Gilbert, 1987; Abramowitz & Stegun, 1964).

We created normal and lognormal probability plots for each age group and each gender. Exhibits 4 and 5 are examples presenting normal and lognormal probability plots for female's and male's body weights 12 to 13 years of age. The points on each graph represent the percentiles reported by the NHANES II Survey, and each straight line represents a least-squares linear regression (Eqs 3 and 4).

From visual inspection of the 80 plots (20 normal and 20 lognormal probability plots for each gender), we observed that the linear regressions on the lognormal probability plots fit the percentile points much better than those on the normal probability plots. To compare the two models quantitatively, we examined several goodness-of-fit measures

for the regressions (D'Agostino & Stephens, 1986). In Exhibits 6 and 7, the t statistics around the $\hat{\mu}_2$ and the $\hat{\sigma}_2$ values indicate that the lognormal distributions consistently fit the points better than do the normal distributions. Also, the adjusted $R^2$ values ( $aR^2$) and the F ratios for the regressions show that the lognormal distributions consistently fit the percentiles better than do the normal distributions. From these graphical and quantitative comparisons, we conclude that the lognormal distributions consistently give strong fits to the percentile of body weights for each gender across all age groups, a finding in turn consistent with results previously published for adult women and men in the United States (Brainard & Burmaster, 1992).

**Lognormal Distributions as a Function of Age** This section in revision

Exhibit 8 graphs the $\hat{\mu}_2$ and the $\hat{\sigma}_2$ values estimated from the lognormal probability plots for female's and male's body weights as a function of age. The top panel shows that the $\hat{\mu}_2$ values for females and males increase relatively smoothly and equally until diverging near age 15 years. The $\hat{\sigma}_2$ values for each gender, however, show much larger relative fluctuations across the age groups, another manifestation of the relatively large sampling errors and statistical fluctuations in the NHANES II results. Note the vertical scales in the two panels are different.

For interpolation and simulations in Mathematica™ (Wolfram, 1991), we fit $n^{th}$-order polynomials to the $\hat{\mu}_2$ and $\hat{\sigma}_2$ values for each gender as a function of age using this functional form:

$$y(t) = c_0 + c_1 \bullet t + c_2 \bullet t^2 + \ldots + c_n \bullet t^n \quad \text{for} \quad 0.5 \le t \le 20 \qquad \text{Eq 5}$$

where y is the fitted variable, t is the age in years, and the n are consecutive integers. We write this polynomial more compactly as the tuple of coefficients, namely $\{c_0, c_1, c_2, \ldots, c_n\}$. For $\hat{\mu}_2$ and $\hat{\sigma}_2$, we fit polynomials in Eq 7 with $2 \le n \le 7$.

Balancing the competing objectives of fidelity of fit, parsimony of expression, and commonality of functional form for females and males, we found that quadratic polynomials gave excellent fits to the $\hat{\mu}_2$ values and that quartic polynomials gave adequate fits to the $\hat{\sigma}_2$ values as a function of age. For females, we found:

$\hat{\mu}_2$ = {2.05065, 0.194101, -0.00447392}, and

$\hat{\sigma}_2$ = {0.164669, -0.0361079, 0.00943697, -0.00070061, 0.0000158666}

with $aR^2$ = 0.995 and 0.791, respectively. For males, we found:

$\hat{\mu}_2$ = {2.17253, 0.164478, -0.00278956}, and

$\hat{\sigma}_2$= {0.165078, -0.0435777, 0.010562, -0.000752062, 0.000016586}

with $aR^2$ = 0.996 and 0.755, respectively.

For each gender, Exhibit 9 presents the fit of the quadratic polynomials superimposed on the point estimates ($\hat{\mu}_2$) and error bars ($\pm\ \hat{\sigma}_2$) for body weight as a function of age.

In this section, we fit polynomial functions of age to the $\hat{\mu}_2$ and the $\hat{\sigma}_2$ values for each gender. These polynomials fit and smooth the time dependencies for the $\hat{\mu}_2$ and the $\hat{\sigma}_2$ values previously fit for each age group. In the appendix, a colleague presents an alternate approach, based on the method of maximum likelihood, that fits the lognormal parameters and the time dependencies in one unified optimization for each gender.

## Practical Application of these Results

In this section, we discuss three applications of these results in probabilistic risk assessments which include children in the exposed population.

First, a risk assessor may include two random generators, one for each gender, for each age group in the study. This direct method has a high computational cost.

Second, a risk assessor may build a custom, parametric generator for a particular problem. To illustrate this approach, we consider a hypothetical situation for which the population consists of an equal number of females and males from 6 months to 7 years of age. Using lognormal distributions with the $\hat{\mu}_2$ and $\hat{\sigma}_2$ values taken from Exhibits 6 and 7, we had Crystal Ball™ simulate 1,000 individuals in each age group for each gender. In Exhibit 10, we pooled these values, plotted them on a lognormal probability plot using Mathematica™, and estimated $\hat{\mu}_2 = 2.69$ and $\hat{\sigma}_2 = 0.326$ for this hypothetical population from the least-squares linear regression ($aR^2 = 0.991$). Exhibit 11 shows the resulting PDF and CDF for this mixed group of children. This method provides an excellent fit for the central 95 percent of the distribution in this example, (e.g., $| z | \leq 2$), but it fails when the population includes a wider age range because the pooled values no longer follow a parametric distribution.

Third, a risk assessor may build two custom generators, one for each gender, based on the polynomial models in this main report or on the alternative models in the appendix. The generators so constructed take the age range of interest as an input. If well designed, these modules can be re-used in many different assessments in the sense of object-oriented programming. In a future report, we will illustrate this most general approach.

## Discussion

Starting for the most recent and best available data for the 50 states, we have found that lognormal distributions give consistently strong fits to the body weights of children, ages 6 months to 20 years. This paper and its appendix present two alternative and equally practical ways to model and then simulate the age dependencies of these lognormal distributions for each gender.

## Acknowledgments

We thank Michaèl Rowland of the National Center for Health Statistics for his help in obtaining the NHANES II Survey data. We also thank William M. Rand of the Department of Community Health at Tufts University for many constructive suggestions.

General Electric Company and Alceon Corporation funded this research.

## Trade Marks

Apple and Macintosh are registered trademarks of Apple Computer, Inc.
Excel™ is a registered trademark of Microsoft Corporation.
Mathematica™ is a registered trademark of Wolfram Research, Inc.
Crystal Ball™ is a registered trademark of Decisioneering, Inc.
Quattro Pro™ is a registered trademark of Borland International, Inc.

## References

Abramowitz & Stegun, 1964
Abramowitz, M. and I.A. Stegun, Eds, 1964, Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables, National Bureau of Standards, Applied Mathematics Series Number 55, Issued June 1964, Tenth Printing with corrections in December 1972, Superintendent of Documents, US Government Printing Office, Washington, DC

Brainard & Burmaster, 1992
Brainard, J. and D.E. Burmaster, 1992, Bivariate Distributions for Height and Weight of Men and Women in the United States, Risk Analysis, Volume 12, Number 2, pp 267 - 275

Burmaster & Harris, 1993
Burmaster, D.E. and R.H. Harris, 1993, The Magnitude of Compounding Conservatisms in Superfund Risk Assessments, Risk Analysis, Volume 13, Number 2, pp 131 - 134

Chambers et al, 1983
Chambers, J.M., W.S. Cleveland, B. Kleiner, and P.A. Tukey, 1983, Graphical Methods for Data Analysis, Wadsworth International Group, Belmont, CA, and Duxbury Press, Boston, MA

Cleveland, 1985
Cleveland, W.S., 1985, The Elements of Graphing Data, Wadsworth Advanced Books, Monterey, CA

D'Agostino & Stephens, 1986
D'Agostino, R.B. and M.A. Stephens, Editors, 1986, Goodness-of-Fit Techniques, Marcel Dekker, New York, NY

Evans et al, 1993
Evans, M., N. Hastings, and J.B. Peacock, 1993, Statistical Distributions; Second Edition, John Wiley & Sons, Inc., New York, NY

Gilbert, 1987
Gilbert, R.O., 1987, Statistical Methods for Environmental Pollution Monitoring, Van Nostrand Reinhold, New York, NY

NAS, 1983
US National Academy of Sciences, 1983, Risk Assessment in the Federal Government: Managing the Process, March 1983

NCHS, 1987
National Center for Health Statistics, Anthropometric Reference Data and Prevalence of Overweight United States, 1976-1980, Data from the National Health Survey Series 11, No. 238, DHHS Publication No. (PHS) 87-1668, US Department of Health and Human Services, Hyattsville, MD, October 1987

Tukey, 1977
Tukey, J.W., 1977, Exploratory Data Analysis, Addison-Wesley, Reading, MA

US EPA , 1989, EFH
US Environmental Protection Agency, 1989, Final Report Exposure Factors Handbook, Office of Health and Environmental Assessment, EPA/600/8-89/043, May 1989

Wolfram, 1991
Wolfram, S., 1991, Mathematica, A System for Doing Mathematics by Computer, Second Edition, Addison-Wesley, Redwood City, CA

## Exhibit 1A

### Female's Body Weights; 6 Months to 20 Years of Age

| Age (years) | Number Examined (n) | Mean (kg) | 5th Percentile (kg) | 10th Percentile (kg) | 15th Percentile (kg) | 25th Percentile (kg) | 50th Percentile (kg) | 75th Percentile (kg) | 85th Percentile (kg) | 90th Percentile (kg) | 95th Percentile (kg) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 months to 1 | 177 | 8.8 | 6.6 | 7.3 | 7.5 | 7.9 | 8.9 | 9.4 | 10.1 | 10.4 | 10.9 |
| 1 to 2 | 336 | 10.8 | 8.9 | 9.1 | 9.4 | 9.9 | 10.7 | 11.7 | 12.4 | 12.7 | 13.4 |
| 2 to 3 | 336 | 13.0 | 10.8 | 11.3 | 11.6 | 12.0 | 12.7 | 13.8 | 14.5 | 14.9 | 15.9 |
| 3 to 4 | 366 | 14.9 | 11.7 | 12.3 | 12.9 | 13.4 | 14.8 | 16.1 | 17.0 | 17.5 | 18.4 |
| 4 to 5 | 396 | 17.0 | 13.8 | 14.3 | 14.6 | 15.2 | 16.7 | 18.4 | 19.3 | 20.2 | 21.2 |
| 5 to 6 | 364 | 19.6 | 15.3 | 16.2 | 16.7 | 17.3 | 19.0 | 21.2 | 22.8 | 24.7 | 26.6 |
| 6 to 7 | 135 | 22.2 | 17.1 | 17.8 | 18.6 | 19.3 | 21.3 | 23.8 | 26.6 | 29.0 | 29.6 |
| 7 to 8 | 157 | 24.7 | 19.2 | 19.5 | 19.9 | 21.5 | 23.8 | 27.1 | 28.7 | 30.3 | 34.1 |
| 8 to 9 | 123 | 27.9 | 21.5 | 22.4 | 23.3 | 24.4 | 27.6 | 30.2 | 31.4 | 33.3 | 36.5 |
| 9 to 10 | 149 | 32.0 | 23.0 | 25.0 | 25.9 | 27.0 | 29.7 | 33.6 | 39.4 | 43.4 | 48.5 |
| 10 to 11 | 136 | 36.1 | 25.8 | 27.5 | 29.1 | 31.0 | 34.5 | 39.5 | 44.3 | 45.9 | 49.7 |
| 11 to 12 | 140 | 41.9 | 29.9 | 30.4 | 31.4 | 34.0 | 40.4 | 45.9 | 51.1 | 56.7 | 60.1 |
| 12 to 13 | 147 | 46.5 | 32.4 | 35.1 | 36.8 | 39.2 | 45.4 | 52.7 | 58.2 | 60.4 | 64.4 |
| 13 to 14 | 162 | 51.0 | 35.5 | 39.1 | 39.5 | 44.2 | 49.1 | 55.3 | 61.0 | 66.6 | 76.4 |
| 14 to 15 | 178 | 54.8 | 40.3 | 42.9 | 43.7 | 47.5 | 53.2 | 60.4 | 65.8 | 67.7 | 75.3 |
| 15 to 16 | 145 | 55.2 | 44.1 | 45.2 | 46.6 | 48.3 | 53.4 | 59.7 | 62.3 | 65.6 | 76.7 |
| 16 to 17 | 170 | 58.1 | 44.2 | 47.4 | 48.9 | 51.3 | 55.7 | 62.6 | 69.0 | 73.4 | 76.9 |
| 17 to 18 | 134 | 59.7 | 44.5 | 48.9 | 50.5 | 52.3 | 58.5 | 63.5 | 68.3 | 71.7 | 81.9 |
| 18 to 19 | 170 | 59.0 | 45.3 | 49.6 | 50.8 | 52.9 | 56.5 | 63.1 | 66.1 | 70.2 | 78.1 |
| 19 to 20 | 158 | 60.2 | 48.6 | 49.8 | 51.8 | 54.0 | 57.2 | 64.5 | 70.7 | 75.0 | 78.2 |
| | 4,079 | | | | | | | | | | |

Source: National Center for Health Statistics: Anthropometric Reference Data and Prevalence of Overweight, US 1976-80
Converted from Pounds to Kilograms

## Exhibit 1B

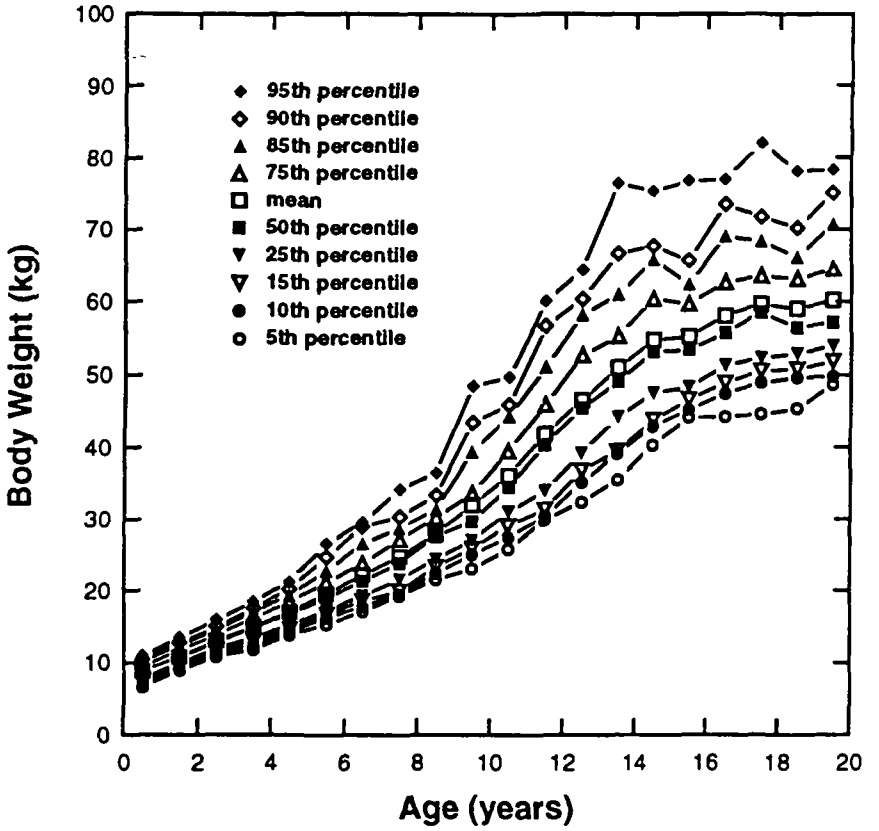
### Male's Body Weights; 6 Months to 20 Years of Age

| Age (years) | Number Examined (n) | Mean (kg) | 5th Percentile (kg) | 10th Percentile (kg) | 15th Percentile (kg) | 25th Percentile (kg) | 50th Percentile (kg) | 75th Percentile (kg) | 85th Percentile (kg) | 90th Percentile (kg) | 95th Percentile (kg) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 months to 1 | 179 | 9.4 | 7.5 | 7.6 | 8.2 | 8.6 | 9.4 | 10.1 | 10.7 | 10.9 | 11.5 |
| 1 to 2 | 370 | 11.8 | 9.7 | 10.0 | 10.4 | 10.8 | 11.7 | 12.6 | 13.2 | 13.6 | 14.4 |
| 2 to 3 | 375 | 13.6 | 11.1 | 11.6 | 11.8 | 12.6 | 13.5 | 14.5 | 15.2 | 15.8 | 16.6 |
| 3 to 4 | 418 | 15.7 | 12.9 | 13.5 | 14.0 | 14.4 | 15.5 | 16.8 | 17.4 | 18.0 | 19.1 |
| 4 to 5 | 404 | 17.8 | 14.1 | 15.0 | 15.3 | 16.0 | 17.6 | 19.1 | 20.0 | 20.9 | 22.2 |
| 5 to 6 | 397 | 19.8 | 16.0 | 16.8 | 17.2 | 17.7 | 19.4 | 21.3 | 22.9 | 23.7 | 25.4 |
| 6 to 7 | 133 | 23.0 | 18.6 | 19.2 | 19.9 | 20.3 | 22.1 | 24.1 | 26.5 | 28.4 | 30.1 |
| 7 to 8 | 148 | 25.1 | 19.7 | 20.8 | 21.2 | 22.2 | 24.9 | 26.9 | 28.3 | 29.6 | 34.0 |
| 8 to 9 | 147 | 28.3 | 20.5 | 22.7 | 23.6 | 24.7 | 27.6 | 30.0 | 33.1 | 35.6 | 39.2 |
| 9 to 10 | 145 | 31.2 | 24.1 | 25.7 | 26.0 | 27.1 | 30.2 | 33.1 | 35.4 | 38.7 | 43.2 |
| 10 to 11 | 157 | 36.5 | 27.2 | 28.3 | 29.6 | 31.5 | 34.9 | 39.3 | 43.6 | 46.3 | 53.5 |
| 11 to 12 | 155 | 40.3 | 26.8 | 28.8 | 31.8 | 33.5 | 37.4 | 46.5 | 52.1 | 57.1 | 61.1 |
| 12 to 13 | 145 | 44.3 | 30.8 | 32.6 | 35.5 | 37.8 | 42.6 | 48.9 | 52.6 | 59.0 | 67.7 |
| 13 to 14 | 173 | 49.9 | 35.4 | 37.0 | 38.4 | 40.1 | 48.5 | 56.3 | 59.9 | 64.3 | 70.1 |
| 14 to 15 | 186 | 57.2 | 41.0 | 44.5 | 46.5 | 49.8 | 56.4 | 63.4 | 66.2 | 69.0 | 77.2 |
| 15 to 16 | 184 | 61.1 | 46.3 | 49.2 | 50.7 | 54.3 | 60.2 | 65.1 | 68.9 | 72.9 | 81.4 |
| 16 to 17 | 178 | 67.2 | 51.5 | 54.4 | 56.1 | 58.7 | 64.5 | 73.8 | 78.2 | 82.3 | 91.3 |
| 17 to 18 | 173 | 66.7 | 50.8 | 53.5 | 54.8 | 58.9 | 65.9 | 72.2 | 76.9 | 82.4 | 89.0 |
| 18 to 19 | 164 | 71.1 | 54.2 | 56.7 | 60.4 | 61.6 | 70.5 | 76.7 | 80.1 | 83.6 | 95.4 |
| 19 to 20 | 148 | 71.8 | 56.0 | 58.0 | 60.7 | 64.0 | 69.6 | 78.0 | 84.4 | 86.9 | 92.3 |
| | 4,379 | | | | | | | | | | |

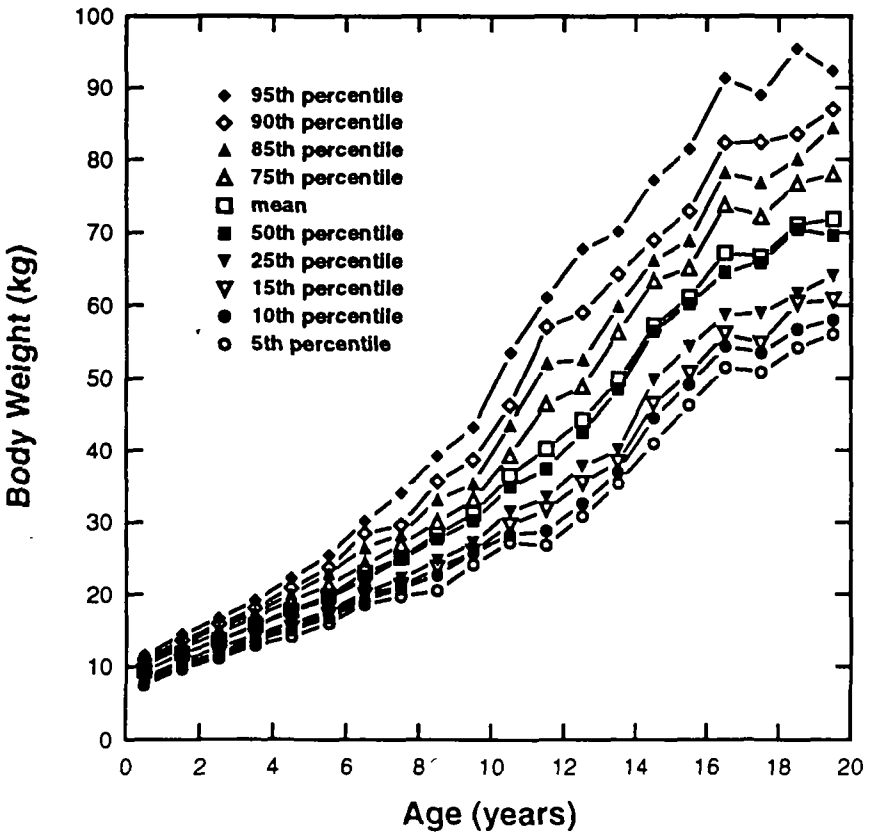
Source: National Center for Health Statistics: Anthropometric Reference Data and Prevalence of Overweight, US 1976-80
Conversion from Pounds to Kilograms

# Exhibit 2

## Female's Body Weights; 6 Months to 20 Years of Age

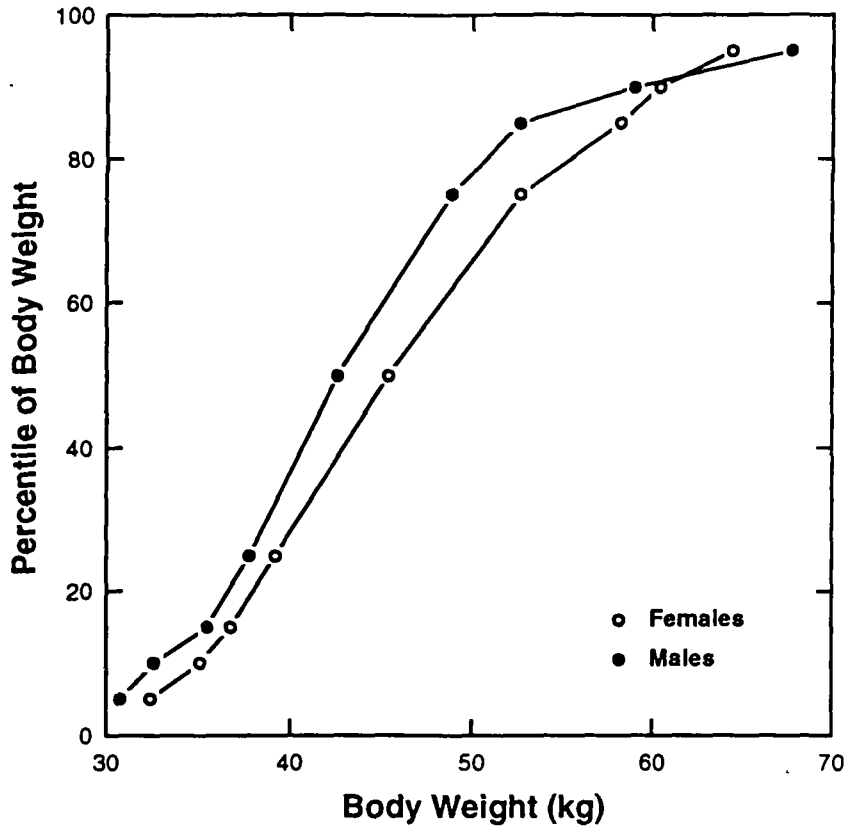



## Male's Body Weights; 6 Months to 20 Years of Age

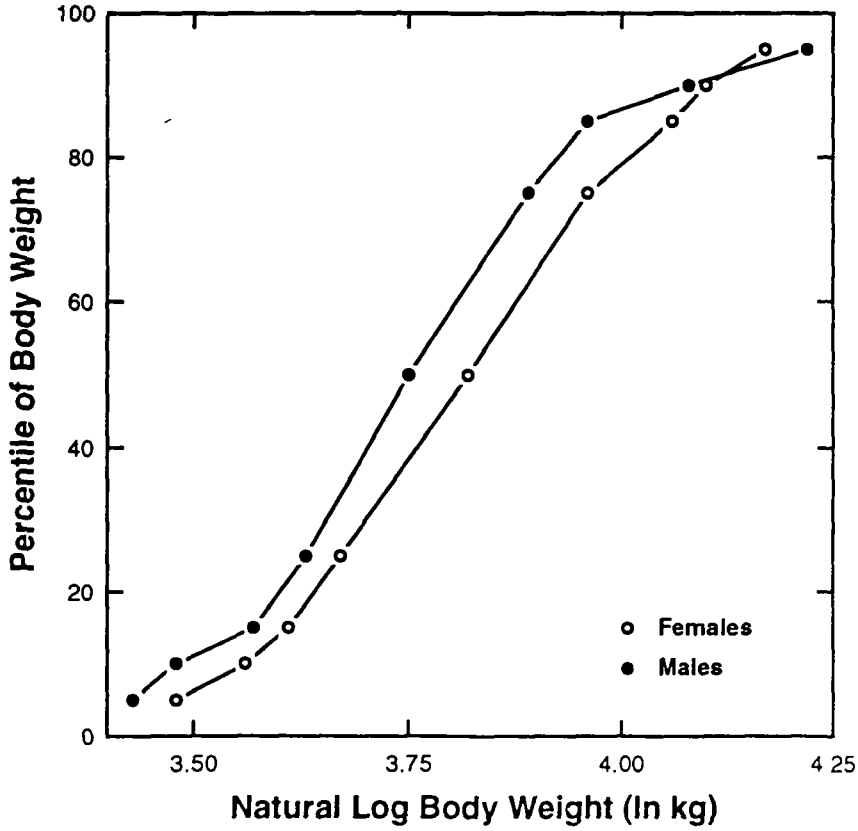


# Exhibit 3

## Cumulative Distribution Function; Females and Males 12 to 13 Years of Age

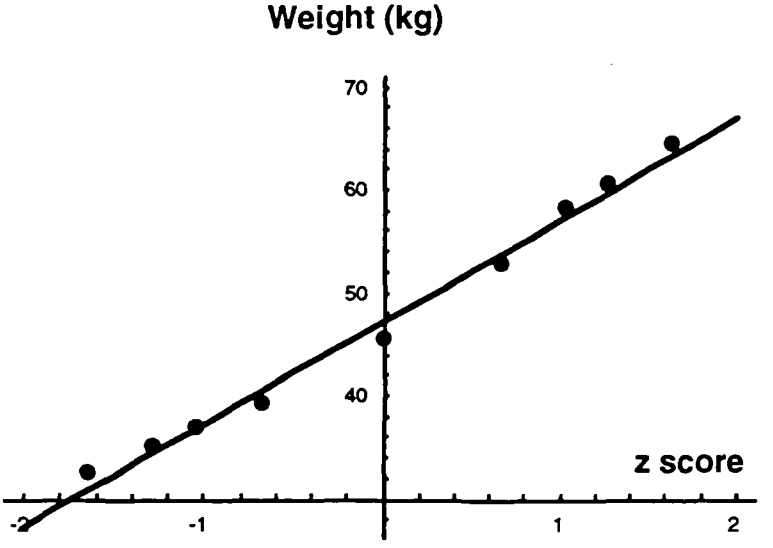



## Cumulative Distribution Function; Females and Males 12 to 13 Years of Age

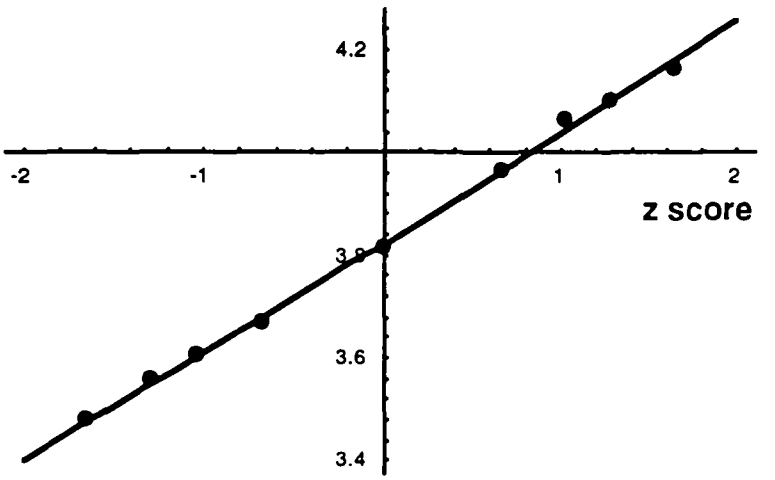


# Exhibit 4

**Normal Probability Plot Female's Body Weights;**
**12 to 13 Years of Age**

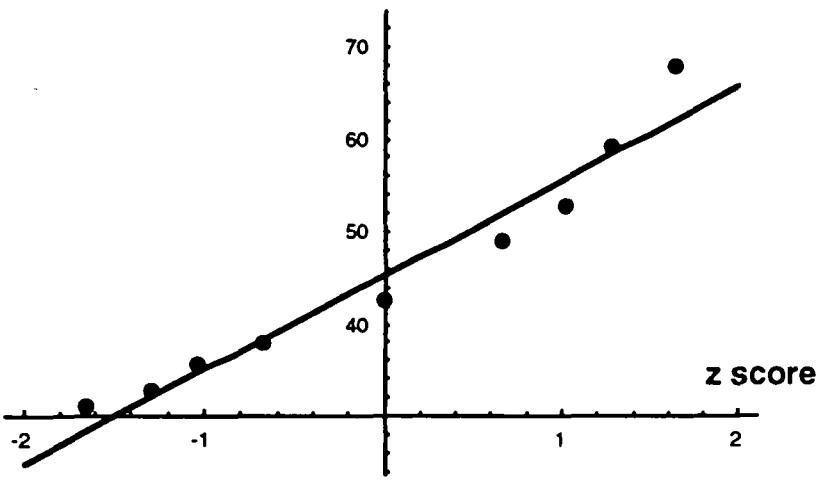



**Lognormal Probability Plot Female's Body Weights;**
**12 to 13 Years of Age**

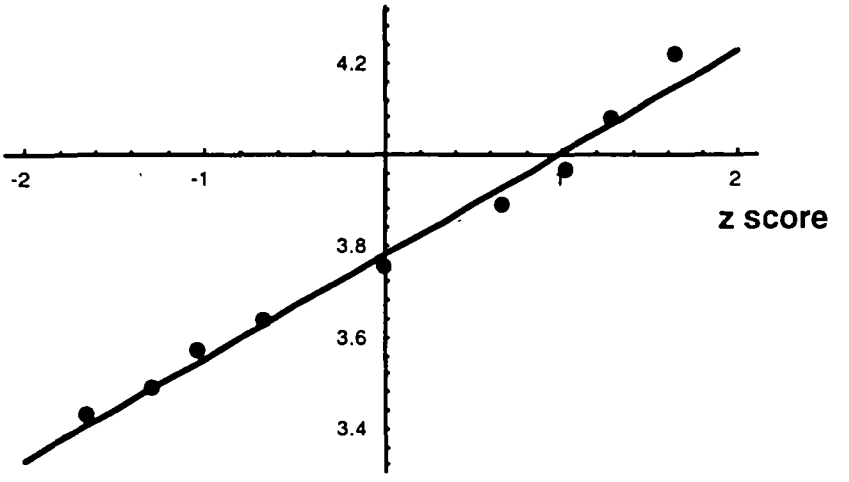


# Exhibit 5

**Normal Probability Plot Male's Body Weights;**
**12 to 13 Years of Age**




**Lognormal Probability Plot Male's Body Weights;**
**12 to 13 Years of Age**



# Exhibit 6
## Statistics for Probability Plot Regression Analyses; Female's Body Weights 6 Months to 20 Years of Age

| Normal Probability Plots | μ1 | tstat | σ1 | tstat | F ratio | aR2 |
|---|---|---|---|---|---|---|
| Age: 6 months to 1 | 8.78 | 219.6 | 1.26 | 35.9 | 1287.3 | 0.994 |
| 1 to 2 | 10.91 | 195.1 | 1.39 | 28.4 | 806.6 | 0.990 |
| 2 to 3 | 13.06 | 158.3 | 1.47 | 20.3 | 412.2 | 0.981 |
| 3 to 4 | 14.90 | 355.2 | 2.02 | 55.1 | 3031.6 | 0.997 |
| 4 to 5 | 17.08 | 155.5 | 2.28 | 23.7 | 561.1 | 0.986 |
| 5 to 6 | 19.98 | 71.1 | 3.27 | 13.3 | 177.0 | 0.957 |
| 6 to 7 | 22.57 | 66.1 | 3.93 | 13.2 | 173.3 | 0.956 |
| 7 to 8 | 24.90 | 65.6 | 4.36 | 13.1 | 172.5 | 0.955 |
| 8 to 9 | 27.84 | 104.2 | 4.33 | 18.5 | 343.4 | 0.977 |
| 9 to 10 | 32.83 | 36.5 | 7.14 | 9.1 | 82.4 | 0.911 |
| 10 to 11 | 36.37 | 86.9 | 7.18 | 19.6 | 384.0 | 0.980 |
| 11 to 12 | 42.21 | 54.8 | 9.51 | 14.1 | 199.3 | 0.961 |
| 12 to 13 | 47.18 | 115.1 | 9.90 | 27.6 | 761.5 | 0.990 |
| 13 to 14 | 51.86 | 47.4 | 11.25 | 11.7 | 138.0 | 0.945 |
| 14 to 15 | 55.20 | 89.9 | 10.29 | 19.2 | 366.9 | 0.979 |
| 15 to 16 | 55.77 | 52.6 | 8.82 | 9.5 | 90.5 | 0.918 |
| 16 to 17 | 58.82 | 85.1 | 9.83 | 16.3 | 264.3 | 0.971 |
| 17 to 18 | 60.01 | 65.2 | 9.93 | 12.3 | 151.8 | 0.950 |
| 18 to 19 | 59.18 | 74.4 | 8.77 | 12.6 | 159.0 | 0.952 |
| 19 to 20 | 61.09 | 77.2 | 9.16 | 13.2 | 175.1 | 0.956 |

| Lognormal Probability Plots | μ2 | tstat | σ2 | tstat | F ratio | aR2 |
|---|---|---|---|---|---|---|
| Age: 6 months to 1 | 2.16 | 343.5 | 0.145 | 26.4 | 695.3 | 0.989 |
| 1 to 2 | 2.38 | 794.5 | 0.128 | 48.7 | 2372.3 | 0.997 |
| 2 to 3 | 2.56 | 618.6 | 0.112 | 30.9 | 955.4 | 0.992 |
| 3 to 4 | 2.69 | 1554.1 | 0.137 | 90.3 | 8147.3 | 0.999 |
| 4 to 5 | 2.83 | 744.1 | 0.133 | 40.1 | 1609.2 | 0.995 |
| 5 to 6 | 2.98 | 327.8 | 0.163 | 20.4 | 418.2 | 0.981 |
| 6 to 7 | 3.10 | 294.2 | 0.174 | 18.8 | 355.3 | 0.978 |
| 7 to 8 | 3.19 | 319.5 | 0.174 | 19.9 | 395.6 | 0.980 |
| 8 to 9 | 3.31 | 585.7 | 0.156 | 31.5 | 992.8 | 0.992 |
| 9 to 10 | 3.46 | 187.2 | 0.214 | 13.2 | 175.2 | 0.956 |
| 10 to 11 | 3.57 | 636.0 | 0.199 | 40.4 | 1636.1 | 0.995 |
| 11 to 12 | 3.71 | 318.0 | 0.226 | 22.1 | 489.6 | 0.984 |
| 12 to 13 | 3.82 | 1102.6 | 0.213 | 70.1 | 4911.1 | 0.998 |
| 13 to 14 | 3.92 | 322.0 | 0.216 | 20.3 | 413.6 | 0.981 |
| 14 to 15 | 3.99 | 732.4 | 0.187 | 39.3 | 1541.9 | 0.995 |
| 15 to 16 | 4.00 | 303.7 | 0.156 | 13.5 | 182.4 | 0.958 |
| 16 to 17 | 4.06 | 539.5 | 0.167 | 25.4 | 644.2 | 0.988 |
| 17 to 18 | 4.08 | 399.4 | 0.165 | 18.5 | 340.7 | 0.977 |
| 18 to 19 | 4.07 | 427.0 | 0.147 | 17.7 | 313.1 | 0.975 |
| 19 to 20 | 4.10 | 427.0 | 0.149 | 17.7 | 314.2 | 0.975 |

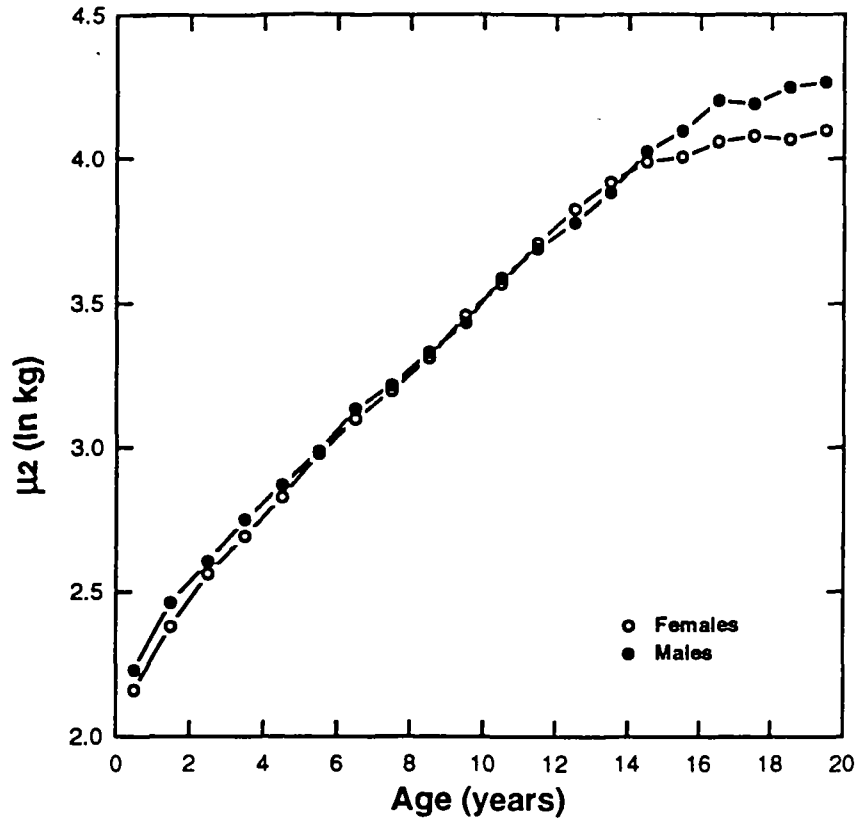
# Exhibit 7
## Statistics for Probability Plot Regression Analyses; Male's Body Weights, 6 Months to 20 Years of Age

| Normal Probability Plots | μ1 | tstat | σ1 | tstat | F ratio | a r2 |
|---|---|---|---|---|---|---|
| Age: 6 months to 1 | 9.39 | 236.2 | 1.23 | 35.2 | 1242.4 | 0.994 |
| 1 to 2 | 11.82 | 231.6 | 1.40 | 31.3 | 982.7 | 0.992 |
| 2 to 3 | 13.63 | 228.6 | 1.64 | 31.4 | 983.1 | 0.992 |
| 3 to 4 | 15.73 | 215.7 | 1.80 | 28.1 | 791.4 | 0.990 |
| 4 to 5 | 17.80 | 184.2 | 2.37 | 28.0 | 784.6 | 0.990 |
| 5 to 6 | 20.04 | 111.9 | 2.78 | 17.7 | 313.6 | 0.975 |
| 6 to 7 | 23.24 | 69.1 | 3.41 | 11.6 | 134.3 | 0.943 |
| 7 to 8 | 25.29 | 66.4 | 3.86 | 11.6 | 134.0 | 0.943 |
| 8 to 9 | 28.56 | 66.6 | 5.17 | 13.8 | 189.3 | 0.959 |
| 9 to 10 | 31.50 | 55.1 | 5.26 | 10.5 | 110.8 | 0.932 |
| 10 to 11 | 37.13 | 47.9 | 7.32 | 10.8 | 116.5 | 0.935 |
| 11 to 12 | 41.68 | 52.1 | 10.42 | 14.9 | 221.7 | 0.965 |
| 12 to 13 | 45.28 | 43.9 | 10.19 | 11.3 | 127.3 | 0.940 |
| 13 to 14 | 50.00 | 74.9 | 10.66 | 18.3 | 333.2 | 0.976 |
| 14 to 15 | 57.11 | 112.5 | 10.25 | 23.1 | 533.2 | 0.985 |
| 15 to 16 | 61.00 | 83.9 | 9.72 | 15.3 | 233.2 | 0.967 |
| 16 to 17 | 67.87 | 78.0 | 11.43 | 15.0 | 225.1 | 0.966 |
| 17 to 18 | 67.16 | 97.9 | 11.21 | 18.7 | 348.5 | 0.977 |
| 18 to 19 | 71.02 | 75.5 | 11.30 | 13.7 | 188.6 | 0.959 |
| 19 to 20 | 72.21 | 133.5 | 11.12 | 23.5 | 552.0 | 0.986 |

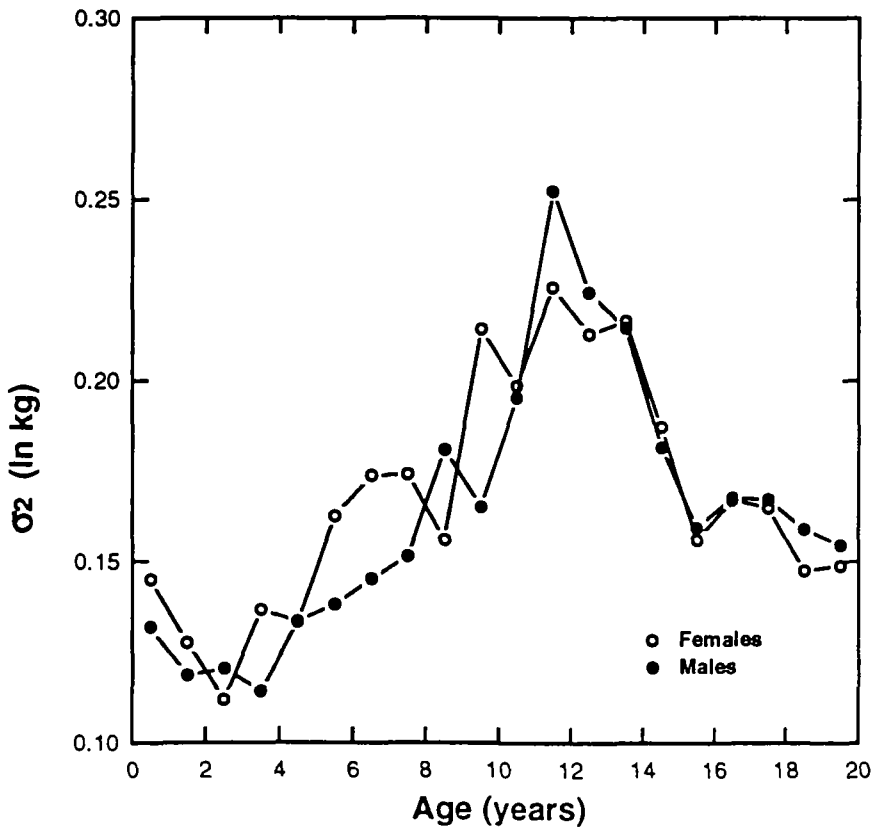
| Lognormal Probability Plots | μ2 | tstat | σ2 | tstat | F ratio | a r2 |
|---|---|---|---|---|---|---|
| Age: 6 months to 1 | 2.23 | 445.2 | 0.132 | 30.1 | 906.0 | 0.991 |
| 1 to 2 | 2.46 | 1133.4 | 0.119 | 62.5 | 3902.3 | 0.998 |
| 2 to 3 | 2.60 | 949.0 | 0.120 | 50.2 | 2516.2 | 0.997 |
| 3 to 4 | 2.75 | 998.7 | 0.114 | 47.5 | 2255.0 | 0.996 |
| 4 to 5 | 2.87 | 1019.2 | 0.133 | 54.2 | 2938.8 | 0.997 |
| 5 to 6 | 2.99 | 524.9 | 0.138 | 27.7 | 769.6 | 0.990 |
| 6 to 7 | 3.13 | 293.3 | 0.145 | 15.5 | 241.3 | 0.968 |
| 7 to 8 | 3.21 | 319.1 | 0.151 | 17.2 | 294.6 | 0.973 |
| 8 to 9 | 3.33 | 343.0 | 0.181 | 21.3 | 452.8 | 0.983 |
| 9 to 10 | 3.43 | 280.0 | 0.165 | 15.4 | 236.7 | 0.967 |
| 10 to 11 | 3.59 | 281.8 | 0.195 | 17.5 | 306.8 | 0.975 |
| 11 to 12 | 3.69 | 341.3 | 0.252 | 26.7 | 712.0 | 0.989 |
| 12 to 13 | 3.78 | 295.6 | 0.224 | 20.1 | 402.3 | 0.980 |
| 13 to 14 | 3.88 | 481.6 | 0.215 | 30.4 | 925.2 | 0.991 |
| 14 to 15 | 4.02 | 764.8 | 0.181 | 39.4 | 1552.4 | 0.995 |
| 15 to 16 | 4.09 | 553.7 | 0.159 | 24 6 | 606.0 | 0.987 |
| 16 to 17 | 4.20 | 557.2 | 0.168 | 25.4 | 646.8 | 0.988 |
| 17 to 18 | 4.19 | 757.3 | 0.167 | 34.5 | 1192.2 | 0.993 |
| 18 to 19 | 4.25 | 479.0 | 0.159 | 20.5 | 419.7 | 0.981 |
| 19 to 20 | 4.26 | 1046.9 | 0.154 | 43.3 | 1877.6 | 0.996 |

# Exhibit 8

## Values of μ2 Estimated from Lognormal Probability Plots

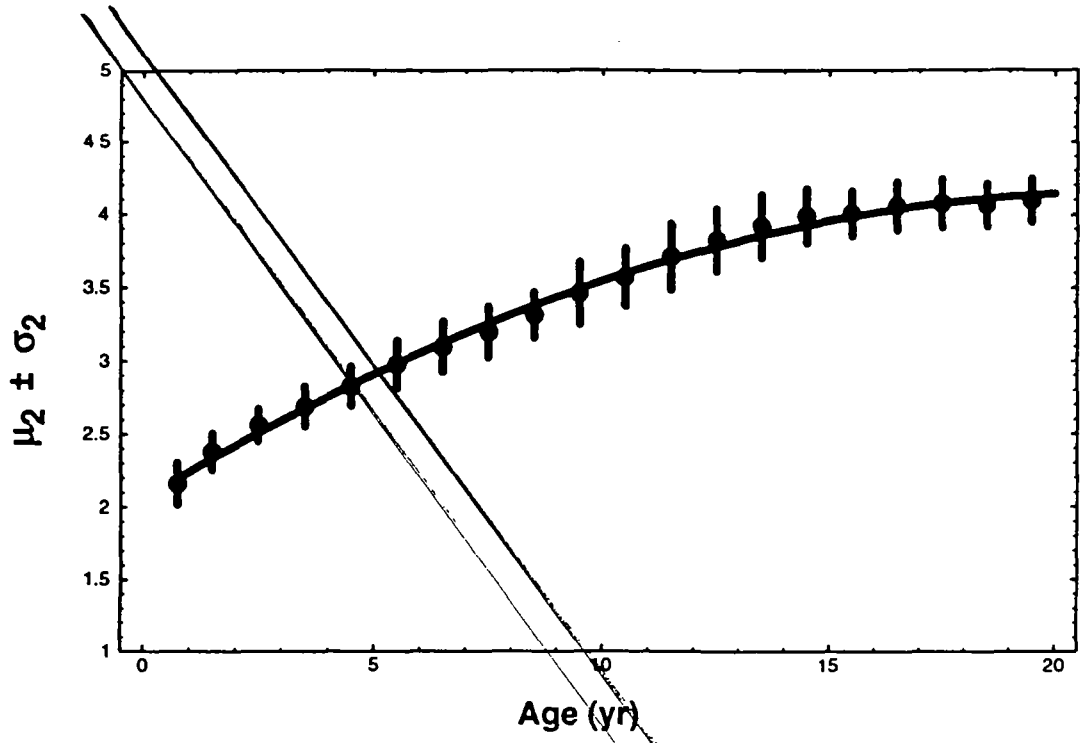



## Values of σ2 Estimated from Lognormal Probability Plots

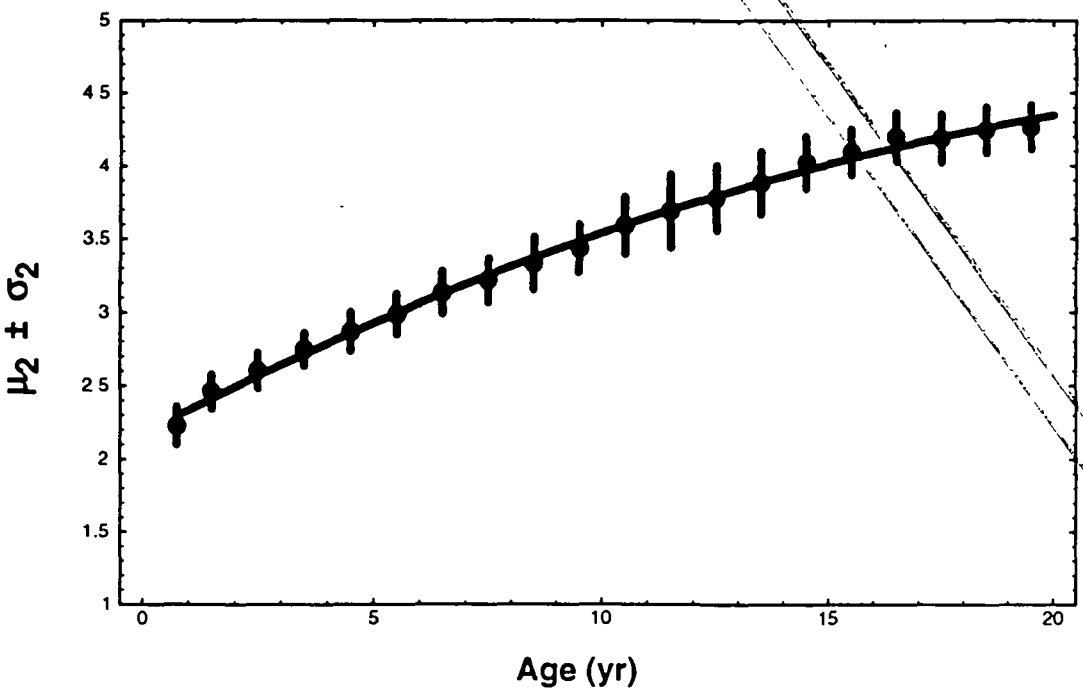


## Exhibit 9

### Polynomial Fit of $\mu_2 \pm \sigma_2$ for Female Body Weight; 6 Months to 20 Years of Age




### Polynomial Fit of $\mu_2 \pm \sigma_2$ for Male Body Weight; 6 Months to 20 Years of Age



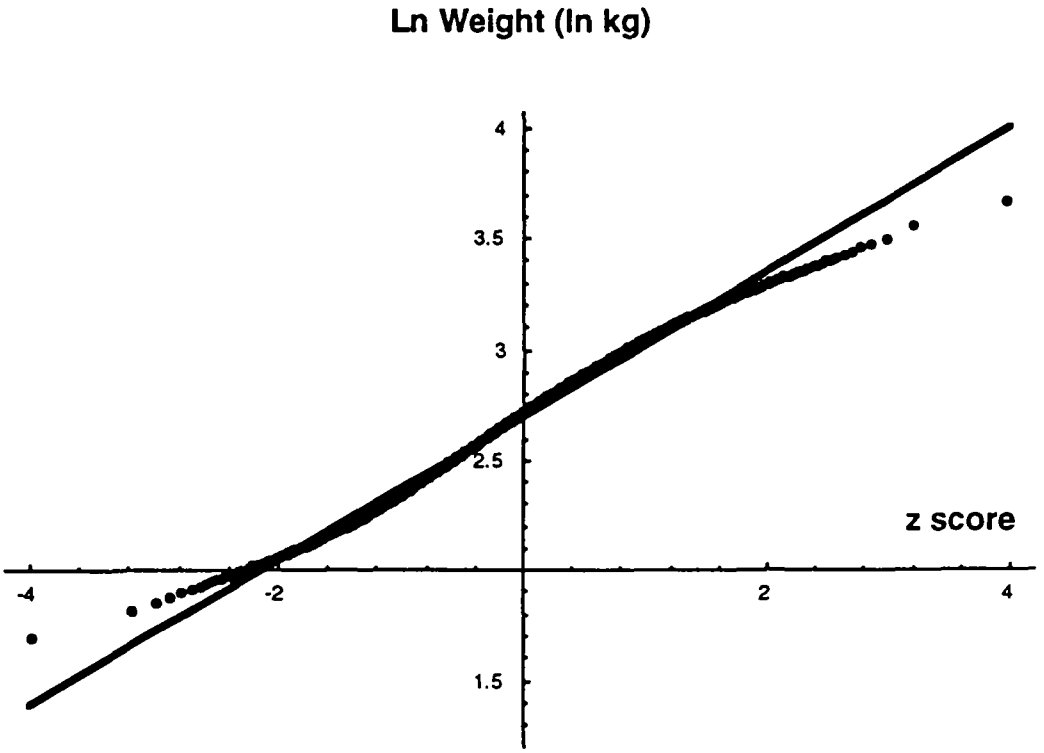

To Be Revised

*Alceon*

# Exhibit 10

## Lognormal Probability Plot for a Mixed Population 6 Months to 7 Years of Age;

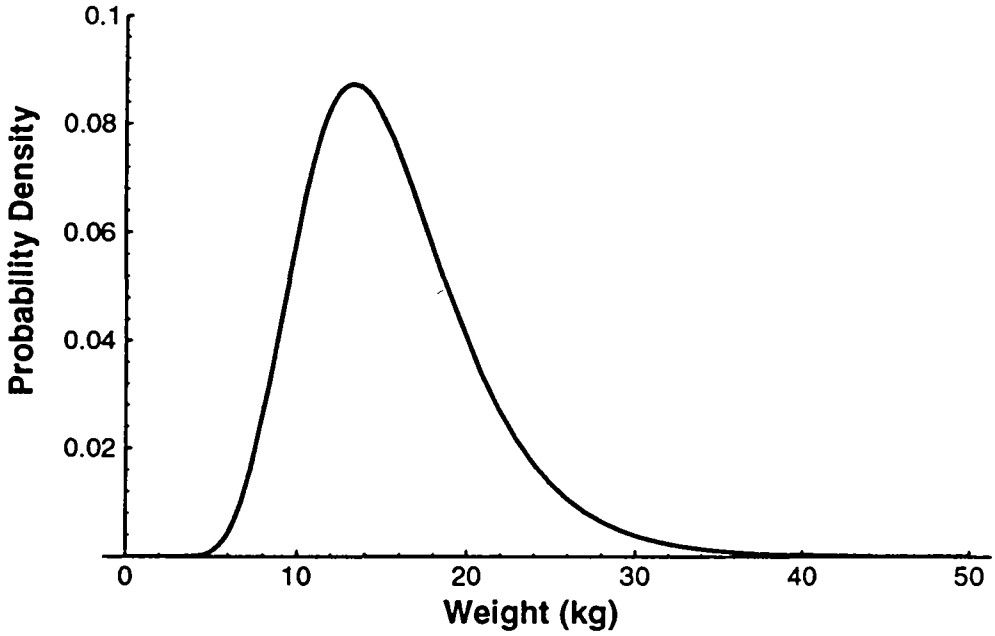
## 1,000 Iterations Per Gender and Age Group Every 10th Point Plotted



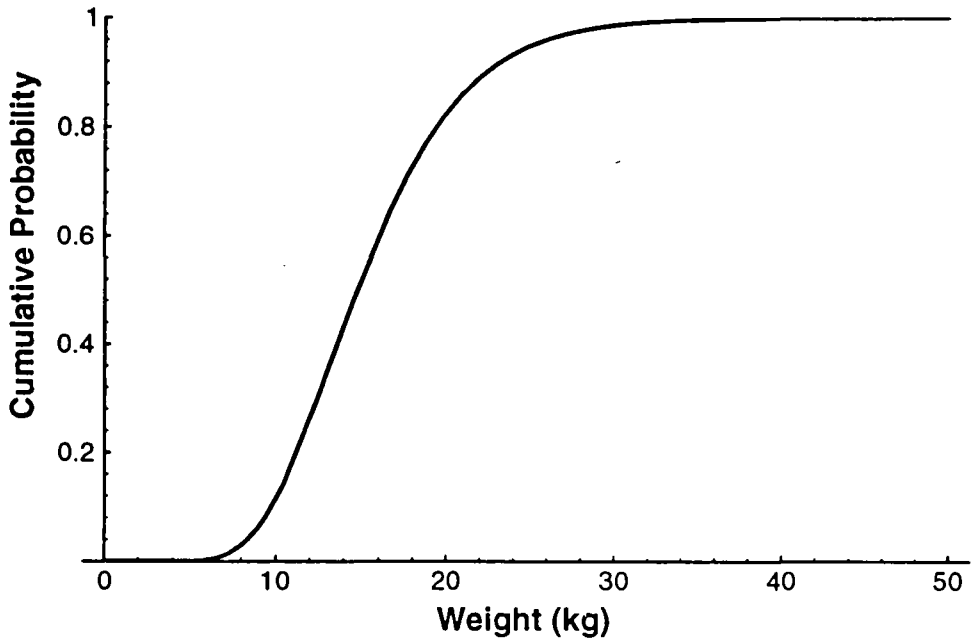

*Alceon*

## Exhibit 11

### Probability Density Function for a Mixed Population; 6 Months to 7 Years of Age




### Cumulative Distribution Function for a Mixed Population; 6 Months to 7 Years of Age




*Alceon*

## Appendix A: A direct approach for interpolation between ages

The main paper provides parameters for linear fits to the transformed empirical cumulative distributions for weight at each age, using inverse normal or inverse lognormal transformation; and it also provides smoothed interpolation-between-age formulae by fitting polynomials to these parameters. For all practical purposes in risk assessment these procedures give excellent results. From a statistical point of view, they may be viewed as sub-optimal because (a) the methods used fail to take account of the correlations between different empirical percentiles, (b) they fail to account for the heteroscedasticity of the empirical percentiles, and (c) the two-stage procedure fails to take account of the heteroscedasticity between ages[1]. We sketch here an extension to the approach that overcomes these objections, can be used to obtain parsimonious and compact expressions to adequately represent all the data shown in Exhibit 1, and may be implemented within a spreadsheet program. As a bonus, we design the interpolation formulae so that they can be extrapolated a few years beyond the age-range given in Exhibit 1 without going seriously astray.

We start with a somewhat more generalized theoretical development than is strictly necessary. Label age groups by $i$, $i = 1, 2, ....., N$, with $n_i$ people measured in age group $i$, with the central

[1] The estimates of percentage points are unbiased, so that the parameter estimates for the distributions are also unbiased. However, they may not be efficient estimates.

age of the age group being $t_i$; and let the $j^{th}$ percentile in age group $i$ be at probability $p_{ij}$, with

$$0 = p_{i0} < p_{i1} < \ldots < p_{iM_i} < p_{iM_i+1} = 1 \qquad \text{Eq. A.1}$$

so that the $M_i$ given percentiles at age $i$ are $p_{i1}$ through $p_{iM_i}$, and these are augmented for convenience with 0 and 1 at each end. Let the value at the $j^{th}$ percentile in age group $i$ be at weight $w_{ij}$, with

$$0 = w_{i0} < w_{i1} < \ldots < w_{iM_i} < w_{iM_i+1} = \infty \qquad \text{Eq. A.2}$$

where again the top and bottom values are appended for convenience of notation. For this example, we transform weights to a logarithmic scale, so the $w$ are natural logarithms of weights, and $w_{i0}$ is taken to be $-\infty$. We further assume that the distribution of ln(weight) at any age is normal, with mean $\mu(t)$ and standard deviation $\sigma(t)$,

If all sampling had been at random from the population, the loglikelihood for the given observations would be:

$$J = \sum_{i=1}^{N} n_i \sum_{j=0}^{M_i} (p_{ij+1} - p_{ij}) \ln \left( \frac{\Phi\left( \frac{w_{ij+1} - \mu(t_i)}{\sigma(t_i)} \right) - \Phi\left( \frac{w_{ij} - \mu(t_i)}{\sigma(t_i)} \right)}{p_{ij+1} - p_{ij}} \right) \qquad \text{Eq. A.3}$$

where $\Phi$ is the standard cumulative normal integral (the denominator $p_{ij+1}-p_{ij}$ in the logarithm has been introduced simply to subtract a constant so that $J$ vanishes for a perfect fit). In fact, the sample was designed as a stratified random sample, and observations were weighted according to the sample design. However, we do not know the sampling weights, nor how

the distributions may differ in the various strata. In the absence of such information, we shall use Eq. A.3 as a suitable approximation to the loglikelihood — and expect that this will give more efficient estimates than the procedure used in the main paper.

All that is now required is a parameterization of the mean $\mu(t)$ and standard deviation $\sigma(t)$, followed by maximization of the likelihood with respect to the parameters to obtain the best fit. The parameterization chosen was:

$$\mu(t) = \ln\left(\frac{(A+Bt)e^{-(t-t_0)/T_0}+D}{1+e^{-(t-t_0)/T_0}}\right) \qquad \text{Eq. A.4}$$

$$\sigma(t) = a+b(t-t_1)\exp\left(-\left(\frac{t-t_2}{T_2}\right)^6\right) \qquad \text{Eq. A.5}$$

The first is a good fit to the general shape of body-weight curves shown in Exhibit 2, and has the advantage of approaching a constant at large ages. The second adequately fits the empirical values of $\sigma$ (see Exhihit A.2 and Exhihit A.3) while also approaching a constant at ages greater than 20.

Maximum likelihood estimates[2] may now be obtained for the parameters. Exhihit A.1 shows these estimates (and the maximum likelihood value of $J$) with sufficient precision to be

---

[2] These estimates were obtained using the optimizer in Borland® Quattro® Pro 5.0 to maximize the loglikelihood given in Eq. A.3. The cumulative normal integral $\Phi$ was implemented using a custom add-in function.

negligibly different from the optimum. These give a very good fit to all the data all at once, as illustrated by comparing the observed mean value with the predicted value, calculated from the predicted median and standard deviation as $\exp(\mu+\sigma^2/2)$. Exhihit A.4 shows the predicted arithmetic mean and standard deviation for these fits, compared with the data points (the standard deviation for the data is computed from the maximum likelihood fits for lognormal distributions at each age independently).

The same approach can also be taken using the polynomial parameterization used in the main paper. Exhihit A.5 shows the mean and standard deviation as calculated using the same order polynomials as given in the main paper, with coefficients selected to maximize the loglikelihood function (Eq. A.3) — these values differ somewhat from those given in the main paper. It can be seen that the parameterization selected in this appendix does considerably better (the maximum likelihood values for the polynomial fits are –298.62 for females, and –300.33 for males).

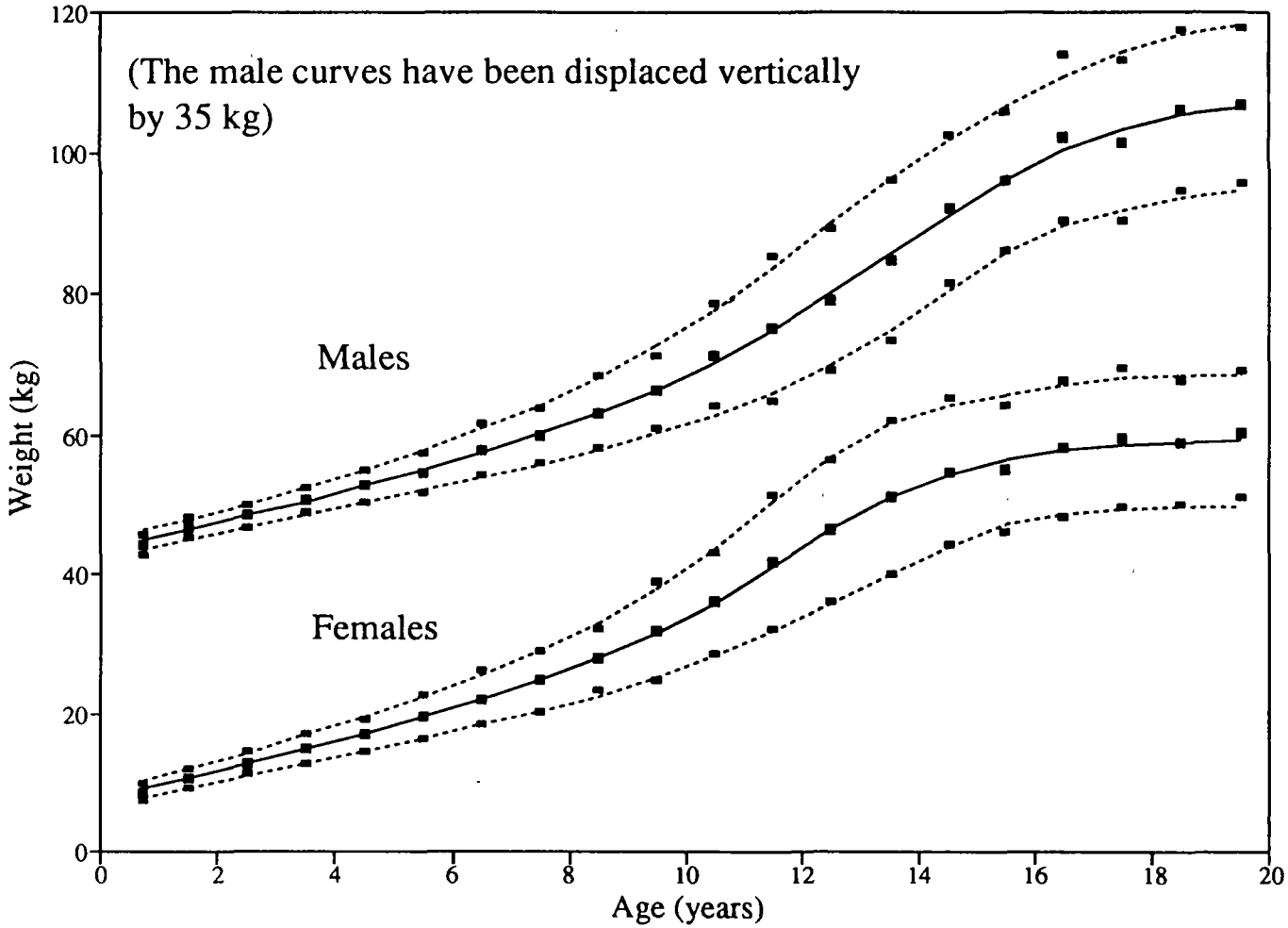
Furthermore, the parameterization of this appendix can be extrapolated a few years beyond the range of the data. At ages greater than 20, both median weight and standard deviation of the weight tend to constant values, and the medians, as estimated from this data, correspond fairly closely to standard estimates of adult body-weight.

Exhihit A.1 Maximum Likelihood Parameters

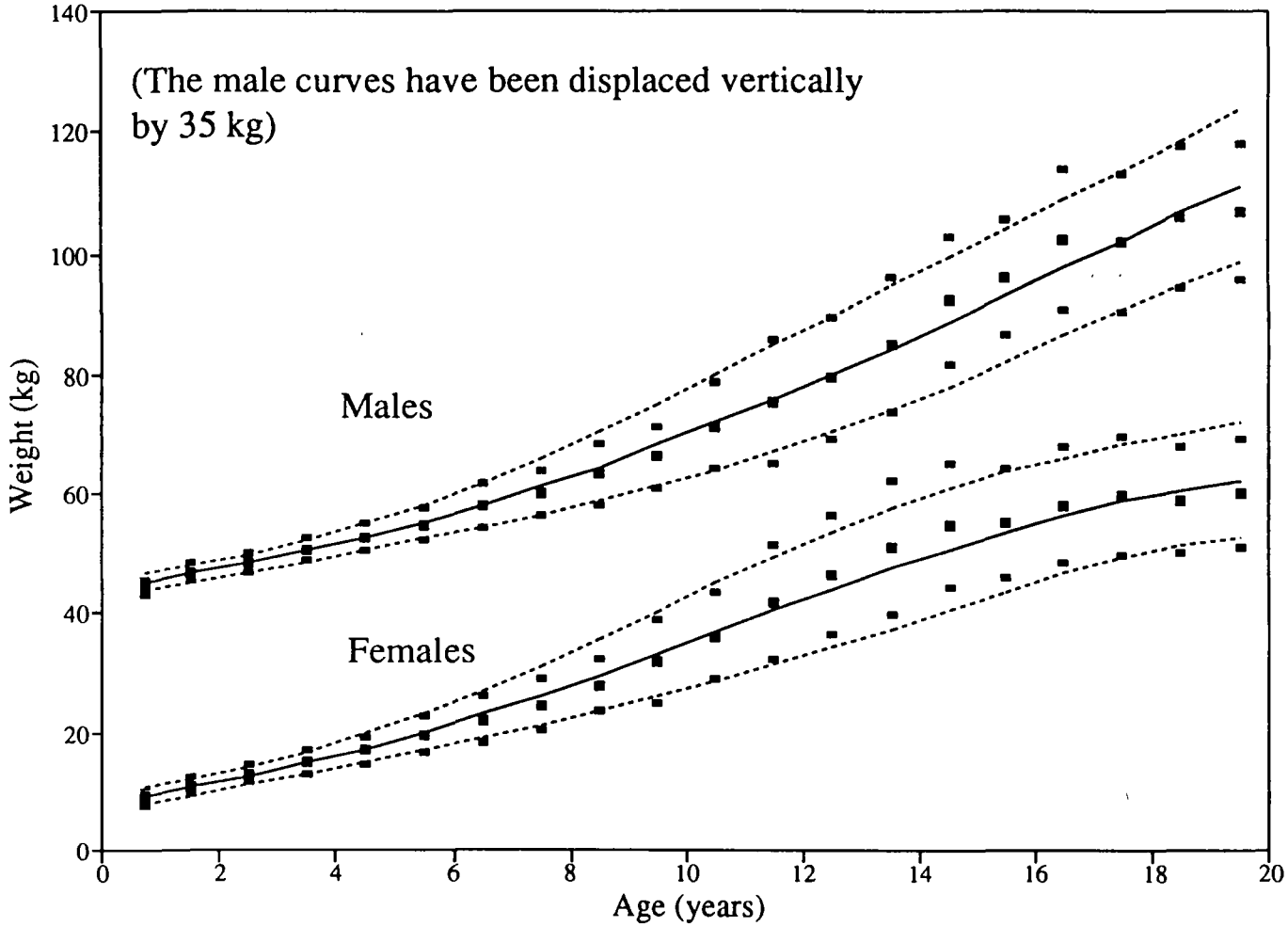
| Parameter | Females | Males | Units |
|---|---|---|---|
| $A$ | 7.430 | 8.270 | kg |
| $B$ | 2.073 | 2.021 | kg/yr |
| $C$ | 58.63 | 72.27 | kg |
| $T_0$ | 1.621 | 1.950 | yr |
| $t_0$ | 12.70 | 14.43 | yr |
| $a$ | 0.1582 | 0.1619 | — |
| $b$ | 0.01205 | 0.01322 | $yr^{-1}$ |
| $t_1$ | 6.000 | 7.031 | yr |
| $T_2$ | 6.582 | 6.530 | yr |
| $t_2$ | 7.593 | 7.650 | yr |
| $J$ | -213.79 | -219.44 | — |


Exhihit A.2 Optimum fit to $\sigma$ for females using the parameterization of Eq. A.5


Exhihit A.3 Optimum fit to σ for males using the parameterization of Eq. A.5




Exhihit A.4 Arithmetic mean and standard deviation using Eq. A.4 and Eq. A.5




Exhihit A.5 Arithmetic mean and standard deviation using polynomial fits

Human and Ecological Risk Assessment Vol 1, No 1, pp 89-100, 1995

# The Need for New Methods to BackCalculate Soil CleanUp Targets in Interval and Probabilistic Cancer Risk Assessments

David E. Burmaster[1], Kylie J. Lloyd[1] and Kimberly M. Thompson[2]

[1]*Alceon Corporation, PO Box 382669, Harvard Square Station, Cambridge, MA 02238-2669*

[2]*Harvard School of Public Health, 665 Huntington Avenue, Boston, MA 02115*

## ABSTRACT

Ordinary algebra may be used to backcalculate health-based cleanup targets in deterministic risk assessments, but it does not work in interval or probabilistic risk assessments. Equations with interval or random variables do not follow the rules of ordinary algebra. This paper explains the need for more sophisticated methods to backcalculate soil cleanup targets when using interval or random variables.

## INTRODUCTION

When estimating incremental lifetime cancer risk, R, associated with environmental exposure to a single carcinogenic chemical via a single exposure pathway, risk assessors often use equations of this fundamental form:

$$R = \frac{\prod_{i=1}^{I} X_i}{\prod_{j=1}^{J} Y_j} \qquad \text{Eqn 1}$$

where $\prod$ indicates a product over the index. In this discussion, we assume that $X_1$ is the exposure point concentration (EPC), $X_I$ is the cancer slope factor (CSF), and all the remaining variables on the right hand side (RHS) of the equation are other exposure variables. Adapting ideas published by the National Academy of Sciences in 1983 (NAS, 1983), the US EPA has published many such equations for use in public health risk assessments at hazardous waste sites (e.g., US EPA, 1989). In all but rare instances, the US EPA has developed its formulae following the general form of Eqn 1 to hold for positive real numbers, i.e., deterministic or point values that do not express either variability or uncertainty in a quantity.

In the deterministic framework, Eqn 3 is always correct, and an analyst can use it to compute the point value for $AccX_1$ consistent with the point value for AccR

## THE INTERVAL PARADIGM

Before discussing the probabilistic paradigm, we first consider the interval paradigm. In this paradigm, each variable no longer takes just a point value but instead takes a range of values within an interval to represent variability and/or uncertainty in a quantity (Alefeld and Herzberger, 1983). An interval variable provides no information on the relative likelihood of any value between the minimum and the maximum. For example, an interval variable $\underline{V}$ that can take any value from 2 to 3, including the endpoints, may be written $2 \leq \underline{V} \leq 3$. We consider *only positive interval variables*, i.e., ones for which both endpoints are positive. We also adopt two new notations for interval variables. First, we underscore each interval variable, e.g., $\underline{V}$, to distinguish one from an ordinary variable. Second, we use a compact notation to show the range, e.g., $\underline{V}$ = [min, max].

### Forward Calculations

In the interval paradigm, Eqn 1 remains the fundamental equation of risk assessment. However, in this framework, the analyst interprets each of the variables on the RHS of Eqn 1 as an interval variable that takes a range of positive values.

To emphasize the change in perspective, we re-write Eqn 1 as Eqn 4 with singly underscored symbols to denote that each variable is now an interval variable:

$$\underline{R} = \frac{\prod_{i=1}^{I} \underline{X}_i}{\prod_{j=1}^{J} \underline{Y}_j} \qquad \text{Eqn 4}$$

In Eqn 4, $\underline{R}$ is an interval variable because each of the $\underline{X}_i$ and $\underline{Y}_j$ is an interval variable.

### Backward Calculations

When working in the interval framework, the risk assessor must solve Eqn 5 for the acceptable exposure point concentration ($\underline{AccX}_1$) that is consistent with the prevailing policy for acceptable risk ($\underline{AccR}$).

$$\underline{AccR} = \frac{\underline{AccX}_1 \cdot \prod_{i=2}^{I} \underline{X}_i}{\prod_{j=1}^{J} \underline{Y}_j} \qquad \text{Eqn 5}$$

Thus Eqn 6′ (and Eqn 6) are false for interval variables. The endpoints of the incorrect interval for $\underline{AccX}_i$ calculated by erroneous Eqn 6′ each differ by a factor of 4 from the correct values in this example. Further, the substitution of the incorrect answer, [1/2, 16] as computed by erroneous Eqn 6′, directly into Eqn 5′ shows that it does not meet the prevailing policy for acceptable risk. Even more unexpected, $\underline{V} / \underline{V} \neq [1, 1]$ in general in the interval paradigm. These results surprise many risk assessors.

This result, well known in mathematics, shows that real number algebra cannot be used to invert Eqn 5 to Eqn 6 for positive interval variables. Under suitable conditions, there are more sophisticated methods which an analyst can use to solve Eqn 5 for positive interval variable $\underline{AccX}_i$ given positive interval variables for the other variables on its LHS and RHS (Alefeld and Herzberger, 1983; Ferson and Long, 1994; Burmaster and Thompson, 1994). Regardless of the method used to calculate a proposed solution for $\underline{AccX}_i$ in Eqn 5, a proposed solution — say, a proposed range for a soil cleanup target — is considered mathematically correct if and only if the proposed solution satisfies Eqn 5 when substituted into it.

## THE PROBABILISTIC PARADIGM

In the rest of this manuscript, we discuss a fully probabilistic paradigm. While a number of authors and the US EPA itself have begun to support a probabilistic interpretation for some of the variables in Eqn 1, we know of no thorough discussion of the consequences of adopting a fully probabilistic interpretation of Eqn 1 when moving from risk assessment to backcalculating cleanup targets as a part of risk management.

Why is it appropriate to replace the deterministic framework with a probabilistic one? We give two of the many answers (for others, see: Morgan and Henrion, 1990). First, in theory, dictionaries base the definition of risk on the concept of chance or probability. For example, the Webster's *New World Dictionary* defines "risk" as "the chance (meaning, probability) of injury, damage, or loss...." (Webster's, 1970). The probabilistic framework returns risk assessment to its most basic definition. Second, as a practical matter, risk assessors agree that all the variables on the RHS of Eqn 1 contain both (i) variability (here, defined as knowledge of heterogeneity in a well-characterized population, usually not reducible through further measurement or study) and/or (ii) uncertainty (here, defined as ignorance about a poorly-characterized phenomenon or model, sometimes reducible through further measurement or study). For example, not every adult drinks the same amount of water each day (a manifestation of variability). Further, an analyst may not know how much water each adult drinks each day (a manifestation of uncertainty). Most mathematicians use random variables to represent and analyze both variability and uncertainty (Morgan and Henrion, 1990; Cooke, 1991), and techniques are available for propagating them independently (if appropriate) using simulation (Frey, 1992; Hoffman and Hammonds, 1993; Carrington, 1993).

Backward Calculations

When interpreting a probabilistic risk assessment, a risk manager also needs a new framework to decide if an acceptable risk or unacceptable risk occurs at a site. She or he can no longer use a simple "bright line test", i.e., a single number, against which to judge the estimated distribution of risk, $\underline{\underline{R}}$. In the probabilistic paradigm, risk is a random variable represented by a probability distribution, so a risk manager must make a decision about the acceptability or unacceptability of risk by making decisions about the acceptability or unacceptability of the *distribution of risk*, not about a single point value of risk. In this manuscript, we investigate $\underline{\underline{AccR}}^*$ as a single specified distribution.

When working in the probabilistic framework, the risk assessor must solve Eqn 7 for the acceptable value of the exposure point concentration ($\underline{\underline{AccX}}_1$) that is consistent with the *distribution* for acceptable risk ($\underline{\underline{AccR}}^*$) specified by the risk manager.

$$\underline{\underline{AccR}}^* = \frac{\underline{\underline{AccX}}_1 \cdot \prod_{i=2}^{I} \underline{\underline{X}}_i}{\prod_{j=1}^{J} \underline{\underline{Y}}_j} \qquad \text{Eqn 8}$$

Risk assessors experienced in using Eqns 1, 2, and 3 in the deterministic paradigm often think that they can use ordinary algebra to re-arrange Eqn 8 into Eqn 9:

$$AccX_1 \neq \frac{\underline{\underline{AccR}}^* \cdot \prod_{j=1}^{J} \underline{\underline{Y}}_j}{\prod_{i=2}^{I} \underline{\underline{X}}_i} \qquad \text{Eqn 9}$$

Unfortunately, in the probabilistic paradigm, Eqn 9 does not follow from Eqn 8 because $\underline{\underline{AccR}}^*$ is not independent, a mistake that we ourselves have made (Lloyd et al., 1992).

To show that Eqn 9 does not follow from Eqn 8, let us exploit a property of lognormal distributions. We use the notation:

$$\underline{\underline{V}} \sim \exp[\, N(\mu_V, \sigma_V)\,]$$

to represent a random variable, $\underline{\underline{V}}$, whose natural logarithm is distributed as a normal or Gaussian random variable with mean $\mu_V$ and standard deviation $\sigma_V$ (Gilbert, 1987).

*Case B*: Eqn 8′ has a degenerate solution for $\underline{\underline{AccX}}_1$ in the form of a real constant

Continuing the numerical example above, let us consider that the risk manager specifies the distribution for acceptable risk as $\underline{\underline{AccR}}^* \sim \exp[N(-1, 5)]$. In this case, we find $\underline{\underline{AccX}}_1 \sim \exp[N(0, 0)]$, a real value, a result confirmed by direct substitution in Eqn 8′. In Case B, $\underline{\underline{AccX}}_1$ degenerates to a constant = $e^0 = 1$ However, Eqn 9′ suggests that $\underline{\underline{AccX}}_1 \sim \exp[N(0, \sqrt{50})]$, a false result that fails direct substitution in Eqn 8′.

*Case C*: Eqn 8′ has no feasible solution for $\underline{\underline{AccX}}_1$.

Continuing the numerical example above, let us consider that the risk manager specifies the distribution for acceptable risk as $\underline{\underline{AccR}}^* \sim \exp[N(-1, 4)]$. In Case C, we find that $\underline{\underline{AccX}}_1$ has no feasible solution in Eqn 8′. However, Eqn 9′ suggests that $\underline{\underline{AccX}}_1 \sim \exp[N(0, \sqrt{41})]$, a false result that fails direct substitution in Eqn 8′.

As the variance of $\underline{\underline{AccR}}^*$ decreases, the solution to Eqn 8′ degenerates from a distribution in Case A, to a real value in Case B, and then to no solution in Case C. In all three cases, Eqn 9′ gives incorrect and misleading results. This conclusion, well known in mathematics, shows that ordinary algebra cannot be used to invert Eqn 8 to Eqn 9 for random variables. Under suitable conditions, there are more sophisticated methods, including a technique called multiplicative deconvolution, which an analyst can use to solve Eqn 8 for random variable $\underline{\underline{AccX}}_1$ given random variables for the other variables on its LHS and RHS (Ferson and Long, 1994). Regardless of the method used to calculate a proposed solution for $\underline{\underline{AccX}}_1$ in Eqn 8, a proposed solution — say, a proposed distribution for a soil cleanup target — is considered mathematically correct if and only if the proposed solution satisfies Eqn 8 when substituted into it.

## DISCUSSION

First, ordinary algebra can be used to re-arrange Eqn 1, the fundamental risk equation for exposure to a single carcinogen in the fully deterministic paradigm, to estimate point values for soil cleanup targets. However, in the interval and the fully probabilistic paradigms, ordinary (real number) algebra cannot be used to re-arrange the fundamental risk equations (Eqns 4 and 7). Under suitable conditions, there are more sophisticated methods which an analyst can use to solve Eqns 5 or 8 for interval variable $\underline{AccX}_1$ or random variable $\underline{\underline{AccX}}_1$ given appropriate information for the other variables on the LHS and the RHS of the particular equation (Ferson and Long, 1994).

Second, a risk assessor and a risk manager must pick a paradigm for a project and then follow the internal logic of that framework throughout both the assessment and the management phases of the project. Each of the three paradigms is internally consistent. Risk assessors or risk managers who begin in one framework and then switch to a different one will inevitably make erroneous calculations and draw erroneous conclusions.

Sixth, we emphasize that not all risk management activities require directly asking the question "How clean is clean enough?". For example, some risk management decisions are structured as choices among a small number of mutually independent alternatives, each of which has an associated technology, cost, and efficacy. We suggest that the techniques of probabilistic risk analysis and decision analysis are very useful in this type of risk management decision, although they are not addressed in this manuscript.

Finally, one of the most pressing issues raised by this manuscript is the need for risk managers to think about how to make judgments on the acceptability or the unacceptability of distributions of risk. As a society, we need ways to pick distributions of acceptable risk — as full distributions — or to identify acceptable distributions of risk — say, as constraints on probability distributions.

## Acknowledgments

Alceon Corporation and the Harvard School of Public Health funded this research. We thank Jeanne W. Appling, Louis Anthony Cox, Jr., Scott Ferson, George M. Gray, and Robert L. Sielken, Jr. for helpful conversations and comments.

## Trademarks

Crystal Ball is a registered trademark of Decisioneering, Inc.
@Risk is a registered trademark of Palisade Corporation.
RiskQ is a registered trademark of Lawrence Livermore Laboratory and the Regents of the University of California.
Mathematica is a registered trademark of Wolfram Research, Inc.

## References

Aitchison, J. and Brown, J.A.C. 1957. *The Lognormal Distribution*, Cambridge, UK, Cambridge University Press.

Alefeld, G. and Herzberger, J. 1983. *Introduction to Interval Computations*, New York, Academic Press

Benjamin, J.R. and Cornell, C. A. *Probability, Statistics, and Decisions for Civil Engineers*, New York, McGraw Hill

Bogen, K. T. 1993. RiskQ An Interactive Approach to Probability, Uncertainty, and Statistics for Use with Mathematica™, Lawrence Livermore Laboratory, Livermore, CA.

Burmaster, D.E. and Thompson, K.M. 1994. BackCalculating CleanUp Targets in Probabilistic Risk Assessments When the Acceptability of Cancer Risk is Defined Under Different Risk Management Policies, *Hum. Ecol. Risk Assess* 1, 101-120.

Carrington, C.D. 1993. *Quantitative Risk and Response*, Contaminants Branch, Center for Food Safety and Applied Nutrition, US Food and Drug Administration, Washington, DC

Cooke, R.M. 1991. *Experts in Uncertainty, Opinion and Subjective Probability in Science*, Oxford, UK, Oxford University Press.

Crow, E.L. and Shumizu, K. 1988. *Lognormal Distributions*, New York, Marcel Dekker.

Decisioneering, Inc., 1994, Manual for Crystal Ball, Denver, CO

Evans, M., Hastings, N. and Peacock, B. 1993. *Statistical Distributions*, Second Edition, New York, John Wiley and Sons.

Human and Ecological Risk Assessment Vol. 1, No 1, pp 101-120, 1995

# BackCalculating CleanUp Targets in Probabilistic Risk Assessments When the Acceptability of Cancer Risk is Defined Under Different Risk Management Policies

**David E. Burmaster[1] and Kimberly M. Thompson[2]**

*[1]Alceon Corporation, PO Box 382669, Harvard Square Station, Cambridge, MA 02238-2669*

*[2]Harvard School of Public Health, 665 Huntington Avenue, Boston, MA 02115*

## ABSTRACT

When evaluating a fully probabilistic risk assessment, say at a hazardous waste site, the risk manager needs a risk management policy that distinguishes an acceptable distribution of risk from an unacceptable one. This manuscript explores several alternative ways to define the acceptability of a distribution of risk. This manuscript also presents methods to backcalculate distributions for cleanup targets under the alternative risk management policies if the need arises.

## INTRODUCTION

When estimating the incremental lifetime cancer risk, R, from an environmental exposure to a single carcinogenic chemical via a single exposure pathway, risk assessors (hereafter, RAs) often use equations of this fundamental form:

$$R = \frac{\prod_{i=1}^{I} X_i}{\prod_{j=1}^{J} Y_j} \qquad \text{Eqn 1}$$

where $\prod$ indicates a product over the index. In this discussion, we assume that $X_1$ is the exposure point concentration (EPC), $X_I$ is the Cancer Slope Factor (CSF), and all the remaining variables on the right hand side (RHS) of the equation are other exposure variables. Adapting ideas published by the National Academy of Sciences in 1983 (NAS, 1983), the US EPA has published many such equations for use in public health risk assessments at hazardous waste sites (e.g., US EPA, 1989). In all but rare instances, the US EPA has developed its formulae in the deterministic paradigm in which all variables on the RHS of Eqn 1 are positive real numbers, i.e.,

## RISK AS A RANDOM VARIABLE

In the probabilistic paradigm, $\underline{\underline{R}}$ in Eqn 2 is a positive random variable represented by a probability distribution because each of the $\underline{\underline{X}}_i$ and $\underline{\underline{Y}}_j$ is a positive random variable represented by a probability distribution. With knowledge of the distributions of all the $\underline{\underline{X}}_i$ and $\underline{\underline{Y}}_j$, an analyst can calculate a closed form expression for the distribution $\underline{\underline{R}}$ in a handful of special cases with independent variables (Springer, 1979). In most practical cases, including those cases with correlated or jointly distributed random variables, the analyst can simulate a numerical approximation to the distribution $\underline{\underline{R}}$ (Rubenstein, 1981; Morgan, 1984). In simulations, the analyst may use (simple) Monte Carlo sampling or (weighted) Latin Hypercube sampling in a simulation program that may run on a typical personal computer or engineering workstation. Many such computer programs are available today, including RiskQ (Bogen, 1993; Wolfram, 1991), Crystal Ball (Decisioneering, 1994), and @Risk (Palisades, 1993).

## THE ACCEPTABILITY OF A DISTRIBUTION OF RISK

When interpreting a probabilistic risk assessment, a risk manager (hereafter, RM) also needs a probabilistic framework in which to decide whether an acceptable risk or unacceptable risk occurs at a site. She or he can no longer use a simple "bright line test," i.e., a single point value (Rosenthal et al., 1992), against which to judge the estimated distribution of risk without picking a moment or a percentile of the distribution. In the probabilistic paradigm, the incremental lifetime cancer risk is a positive random variable represented by a probability distribution. Hence, a RM must make a decision about the acceptability or unacceptability of the risk by making decisions about the acceptability or unacceptability of the distribution of the risk.

In this manuscript, we investigate several possible tests that a risk manager could use to judge the acceptability or unacceptability of the distribution of the risk. We will discuss several approaches (of the innumerable universe of approaches) that a RM could use to judge the acceptability or unacceptability of distribution $\underline{\underline{R}}$ estimated using Eqn 2. For example, a risk manager could use one of these approaches when interpreting a risk assessment at a hazardous waste site to determine if the current conditions (either before or after some remediation) are acceptable or not. If, for current conditions, the distribution $\underline{\underline{R}}$ estimated using Eqn 2 is judged acceptable according to the governing policy, then the RM may conclude that the site needs no (further) remediation or management. On the other hand, if, for the same conditions, the estimated distribution $\underline{\underline{R}}$ is judged unacceptable, then the RM may conclude that the site needs (further) remediation or management. Of course, the RM may also consider other issues (e.g., cost, engineering feasibility, and public acceptance) in the decision.

In this manuscript, we denote the set of all distributions of risk that meet the governing policy for the acceptability of a distribution of risk with the symbol $\underline{\underline{\text{AccR}}}$. In the first approach discussed next, the set $\underline{\underline{\text{AccR}}}$ includes only one distribution which is acceptable under the governing risk management policy, but in each of the remaining approaches, the set $\underline{\underline{\text{AccR}}}$ includes an infinite number of distributions which are acceptable.

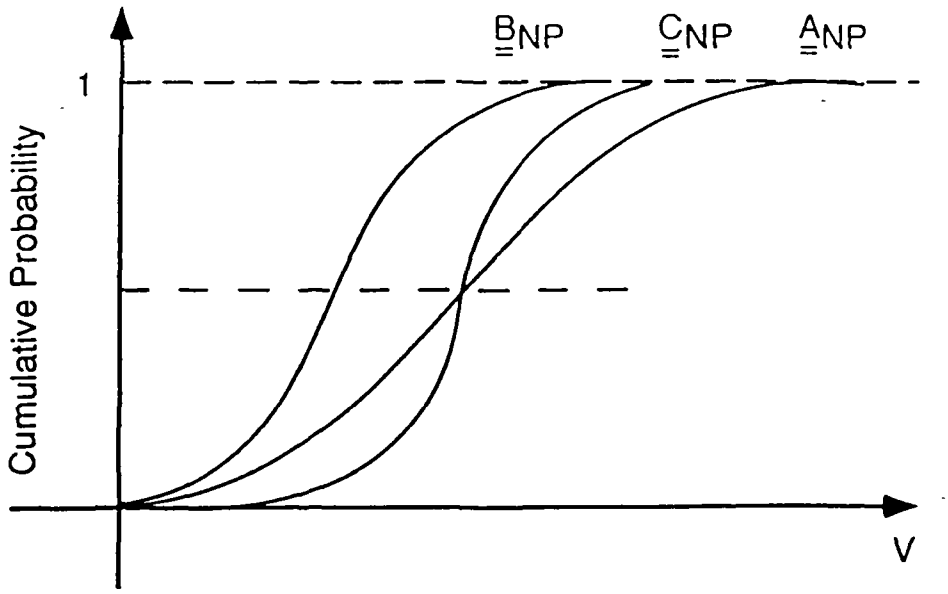



Figure 1. Cumulative distribution functions for three nonparametric distributions.

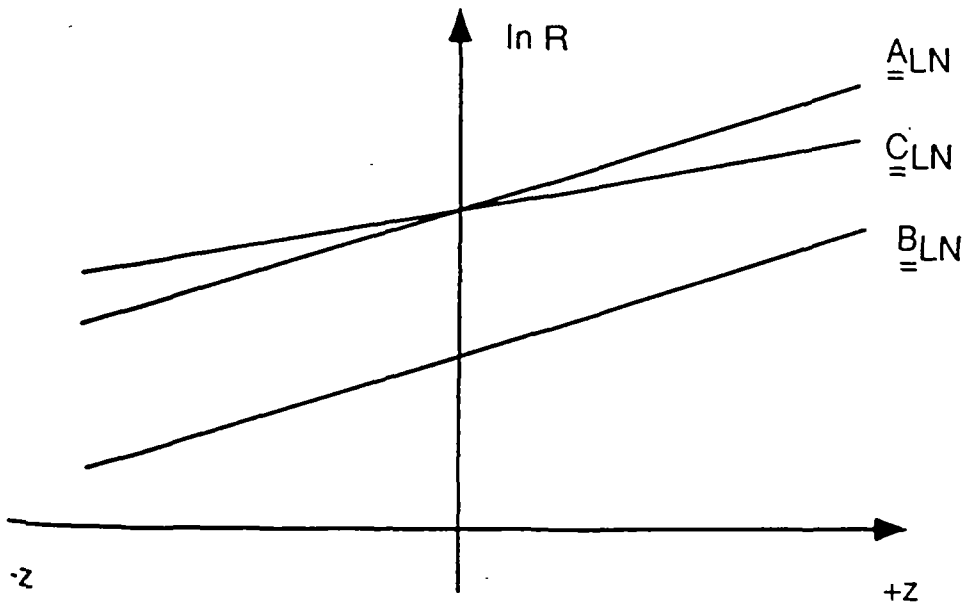



Figure 2. Lognormal probability plots for three lognormal distributions.

### Specifying the Set of Acceptable Distributions using One Inequality Constraint on One Percentile of $\underline{\underline{R}}$

In a fifth and simplest policy approach, the RM could define a distribution of risk $\underline{\underline{R}}$ calculated with Eqn 2 as acceptable as long as one key percentile of the distribution meets one inequality constraint, e.g., Cons 5. Some government agencies have already begun to express their exposure and/or risk management policies for hazardous waste sites along these lines. For example, in Massachusetts, before or after cleanup, a hazardous waste site poses an acceptable risk if the 95$^{th}$ percentile of estimated incremental lifetime cancer risk falls at or below 1 in 100,000 (MA DEP, 1993). In this fifth and simplest policy approach, there are fewer choices for the regulatory agency: (i) the percentile at which the constraint applies, and (ii) the value of the constraint ($\gamma_1$).

## THE ACCEPTABILITY OF A DISTRIBUTION FOR THE EXPOSURE POINT CONCENTRATION

Next, we present ways to backcalculate distributions for the cleanup target under the alternative policy options discussed above.

With the key concept of an acceptable distribution of risk defined as one of the five alternative risk management policies discussed above, one can understand the concept of an acceptable distribution of exposure point concentration as any probability distribution $\underline{\underline{X}}_1$ (or in the degenerate limit of zero variance, any point value $X_1$) which — when substituted into Eqn 2 along with the distributions for the other input variables — yields a distribution $\underline{\underline{R}}$ that meets all criteria for the acceptability of risk specified by the regulatory agency.

If the governing policy for the acceptability of a distribution $\underline{\underline{R}}$ admits an infinite number of distributions into the set $\underline{\underline{AccR}}$ , then, in general, that governing policy in turn will admit an infinite number of distributions of exposure point concentration $\underline{\underline{X}}_1$ into the set of all such acceptable distributions (the set $\underline{\underline{AccX}}_1$).

When conducting a baseline risk assessment (say, before remediation at a hazardous waste site), the RM can decide whether the distribution of exposure point concentration $\underline{\underline{X}}_1$ at a site is acceptable or not by considering the distribution $\underline{\underline{R}}$ using Eqn 2. If the distribution $\underline{\underline{R}}$ is acceptable (or not) according the prevailing regulatory policy, then the distribution $\underline{\underline{X}}_1$ is acceptable (or not). If the distribution $\underline{\underline{R}}$ is not acceptable to the RM according to the governing policy, then some intervention is necessary (i) to make the distribution $\underline{\underline{X}}_1$ "smaller" or narrower, or (ii) to reduce the intensity, frequency, or duration of exposure at the site.

In the fully probabilistic paradigm, the question "How clean is clean enough?" occurs as it does in the fully deterministic paradigm. However, it is usually more difficult to calculate a full distribution for a cleanup target in the fully probabilistic paradigm than it is to calculate a point value for a cleanup target for the exposure point concentration in the deterministic paradigm.

Even though there may be many ways to backcalculate one or more members of the set $\underline{\underline{AccX}}_1$, and even though it may be more difficult to compute one or more members of the set $\underline{\underline{AccX}}_1$, it is straightforward to test a distribution (or a point

### General Solution When AccR* is Specified

In the first policy approach, if the RM specifies one distribution AccR* as the uniquely acceptable distribution of risk, then Eqn 2 has zero or one solution for an acceptable distribution of EPC (Ferson and Long, 1994). If the variance of the specified AccR* is too small (compared to the variances of the other variables), then Eqn 2 has no solution. If the variance of the specified AccR* is large enough, and if other mathematical conditions hold, then Eqn 2 has one solution which may be calculated using multiplicative deconvolution (Ferson and Long, 1994). The single solution may be a single distribution for the exposure point concentration, or, in the limit of decreasing variance, a single real number.

As noted earlier, this approach with one specified AccR* has little or no practical appeal, say, as a way to plan remediation at a hazardous waste site. In addition to the limitation mentioned earlier, this approach has the further limitation that the algorithms for multiplicative deconvolution necessary to compute it are numerically intensive and are sensitive to numerical instabilities (Ferson and Long, 1994).

### General Solution When Either AccR# or AccR+ is Specified

In the second policy approach, if the RM specifies the policy for the acceptability of risk in terms of a fully or partially dominant distribution (AccR# or AccR+, respectively), Eqn 2 has an infinite number of nondegenerate solutions. In this second policy approach, the RA may use a combination of multiplicative deconvolution, dispersive Monte Carlo simulation, and dependency bounds analysis (Ferson and Long, 1994; Ferson, 1994) — along with the methods in Appendix 1 — to find a first solution to the stated problem. Once a first solution is found, the analyst can use numerical experiments and heuristic search to find other solutions closer to the extremal solution. In this second policy approach, the extremal solution is the (unique) feasible solution for $\underline{X}_1$ that yields the R closest to AccR# or AccR+ in a defined metric. If either AccR# or AccR+ is specified, any distribution that is dominated by a known solution to Eqn 2 is also a solution.

### General Solution When Inequality Constraints Are Specified For Moments

In the third policy approach, if the RM specifies the policy for the acceptability of risk in terms of one or more constraints on one or more moments in the form of Cons 1, 2, 3, and/or 4, Eqn 2 always has an infinite number of nondegenerate solutions (unless the constraints somehow contradict each other and admit no feasible solution). In this third policy approach, the RA may use a combination of multiplicative deconvolution, dispersive Monte Carlo simulation, and dependency bounds analysis (Ferson and Long, 1994; Ferson, 1994) — along with the methods in Appendix 1 — to find a first solution to the stated problem. Again, once a first solution is found, the analyst can use numerical experiments and heuristic search to find solutions closer to the extremal solution. In this third policy approach, an extremal solution is any (non unique) nondegenerate distribution that just touches the specified constraints. With the acceptability of risk defined in terms of constraints in the form of Cons 1, 2, 3, and/or 4, any distribution that is dominated by a known solution to Eqn 2 is also a solution.

$$s \cdot \underline{\underline{R}} = s \cdot \underline{\underline{X}}_1 \cdot \frac{\prod_{i=2}^{I} \underline{\underline{X}}_i}{\prod_{j=1}^{J} \underline{\underline{Y}}_j} \qquad \text{Eqn 4}$$

*Discussion*

When inserted into Eqn 2, the scaled distribution $s \cdot \underline{\underline{X}}_1$ leads to the scaled distribution $s \cdot \underline{\underline{R}}$ which, by construction, just touches the constraint(s) that caused the Minimum [ $s_k$ ]. If $\underline{\underline{R}}$ is a lognormal distribution, then $s \cdot \underline{\underline{R}}$ is also a lognormal distribution. Similarly, if $\underline{\underline{X}}_1$ is a lognormal distribution, then $s \cdot \underline{\underline{X}}_1$ is also a lognormal distribution. Finally, any distribution dominated by the distribution $s \cdot \underline{\underline{X}}_1$ is also a solution to the problem.

## General Solution When One Inequality Constraint Is Specified for One Percentile

In the fifth and simplest policy approach, if the RM specifies the policy for the acceptability of risk in terms of one constraint on one specified percentile of the risk distribution in the form of Cons 5, Eqn 2 always has an infinite number of nondegenerate solutions. Again, with the acceptability of risk defined in terms of constraints on one percentile of risk in the form of Cons 5, any distribution that is dominated by a known solution to Eqn 2 is also a solution.

In this policy approach with only one inequality constraint, the scaling method does find one of the infinite number of nondegenerate distributions which are extremal solutions. In this fifth policy approach, an extremal solution is any (non unique) distribution that just meets the single specified constraint at the edge of the feasible envelope. Again, with the acceptability of risk defined in terms of one constraint on one percentile of risk in the form of Cons 5, any distribution that is dominated by a known solution to Eqn 2 is also a solution.

In this fifth policy approach, ChemRisk (1994), Sielken (1994), and McKone (1994) have also found a point value for $X_1$ that is an extremal solution. As an example of this degenerate case, say the single constraint occurs at the 95$^{th}$ percentile of $\underline{\underline{R}}$ as in Cons 5 earlier: $0 < \underline{\underline{R}}_{0.95} \leq \gamma_1$.

If the single inequality occurs at the n$^{th}$ percentile (n > 50), the analyst computes a point value for $X_1$ by taking the (100 - n)$^{th}$ percentile on the RHS of Eqn 5. In this example with the constraint at the 95$^{th}$ percentile of risk, the analyst computes $X_1$ as:

$$X_1 = \left[ \frac{\gamma_1 \cdot \prod_{j=1}^{J} \underline{\underline{Y}}_j}{\prod_{i=2}^{I} \underline{\underline{X}}_i} \right]_{0.05} \qquad \text{Eqn 5}$$

This method also has modest computational burden, and it produces an extremal point value as the cleanup target. In other words, when the point value $X_1$ is substituted into Eqn 2, the 95$^{th}$ percentile (in this example) of the *distribution* $\underline{\underline{R}}$ equals $\gamma_1$. The algebraic proof of this derivation in the special case when all the distributions are lognormal ones shows that the distribution $\underline{\underline{R}}$ that results from this procedure has moments and percentiles which come from the distributions for the $\underline{\underline{X}}_i$ (i = 2, ..., I) and the $\underline{\underline{Y}}_j$ (j = 1, ... J), not from regulatory policy. In parallel with earlier results, any point value (or full distribution) which is stochastically dominated by a known solution is also a solution.

acceptable distribution of risk in terms of inequality constraints on specified percentiles — has only a modest computational burden. Although the scaling algorithm is not an extremal algorithm when two or more inequality constraints are specified as in the fourth policy approach, we find it practical for use in estimating distributions for cleanup targets at many hazardous waste sites, especially ones without prominent "hotspots".

Fifth, although we are not ourselves RMs, we believe that the fourth policy approach — with two or three constraints — is the most attractive of these because it gives the RM great flexibility in specifying a risk management policy in terms of the median risk and the "high end" risk [EndNote 6]. The fourth policy approach also gives the RA certain mathematical methods with modest computational complexity, and the fourth policy approach gives the potentially responsible parties (PRPs) good latitude to fashion a cost-effective remedy consistent with the stated risk management policy. Overall, we believe that the fourth policy approach (not the fifth one) is the most practical and reasonable to pursue in the real world. [EndNote 7]

Sixth, a proposed solution for a nondegenerate distribution (or a proposed point value) of the exposure point concentration — no matter how calculated — *must be verified or falsified* by direct substitution into Eqn 2. Thus, the RM need not understand the method by which someone proposed a full distribution (or point value) for a cleanup target. However, the RM should verify that the proposed distribution $\underline{\underline{X}}_1$ (or the proposed point value $X_1$) is indeed a solution by using the direct test given above.

Seventh, in the fully probabilistic paradigm in which the policy for the acceptability of a distribution of risk is expressed in terms of inequality constraints on moments and/or percentiles, we understand that an engineer working on the remediation at a hazardous waste site must translate the distribution selected by the RM as the cleanup target into explicit instructions for the field crew. [EndNote 4] In this manuscript, we do not consider "where to drive the bulldozer" or any compliance issues. We note that there is no 1:1 relationship between a *probability* distribution for the exposure point concentration and the *spatial* distribution of exposure. [EndNote 5]

Eighth, as a practical matter, RAs and RMs may find themselves working in a hybrid paradigm in which some variables are treated as real numbers (constants) and other variables as random variables. Such a situation may arise for technical or policy reasons. As an example of the former, the RA may decide that it is unnecessary or inappropriate to treat one or several variables as random variables after completing sensitivity analyses of the exposure model. As an example of the latter, the US EPA currently rejects the idea that Cancer Slope Factors are properly modeled as random variables, even though toxicologists inside and outside the Agency recognize that CSFs have both variability and uncertainty inherent in them. In such situations, the RA would use (i) a point value — perhaps an arithmetic mean or a value at a higher percentile — for each of these variables held constant and (ii) a random variable for each of the others. In this hybrid paradigm, an equation similar to Eqn 2 would hold as the fundamental risk equation. Although we do not elucidate the rules for the internal consistency of such hybrid paradigms here, the RA and the RM must do so

5. Many different *spatial* distributions (e.g., for contaminants in soils) may give rise to identical (or indistinguishable) *probability* distributions for the exposure point concentration in a particular risk assessment.
6. By constraining both the median risk and the "high end" risk, the risk management policy indirectly constrains the expected value of risk as well. If the risk management policy only constrains one percentile of risk, even a "high end" percentile of risk, the policy does not limit the expected value of risk, even indirectly.
7. As Albert Einstein wrote, "Make things as simple as possible, but no more so."

## Acknowledgments

We thank Scott Ferson of Applied Biomathematics and Louis Anthony Cox, Jr. of Cox Associates for many helpful conversations and suggestions in formulating and resolving the issues in this manuscript. We also thank Paul S. Price of Chem Risk, Robert L. Sielken, Jr., of Sielken, Inc., Thomas E. McKone of Lawrence Livermore Laboratory, and Mark Harris of Maxus Energy Corporation for providing key information for sections concerning policy options that specify one inequality constraint for one percentile of $\underline{\underline{R}}$.

Alceon Corporation funded DEB's research.

## Trademarks

Crystal Ball is a registered trademark of Decisioneering, Inc. @Risk is a registered trademark of Palisade Corporation. RiskQ is a registered trademark of Lawrence Livermore Laboratory and the Regents of the University of California. Mathematica is a registered trademark of Wolfram Research, Inc.

## References

Alefeld, G. and Herzberger, J. 1983 *Introduction to Interval Computations*, New York, Academic Press

Benjamin, J.R. and Cornell, C.A. 1970. *Probability, Statistics, and Decisions for Civil Engineers*, New York, McGraw Hill.

Bogen, K.T. 1993. RiskQ: An Interactive Approach to Probability, Uncertainty, and Statistics for Use with Mathematica™, Lawrence Livermore Laboratory, Livermore, CA

Burmaster, D.E., Lloyd, K.J. and Thompson, K.M. 1994. The need for new methods to backcalculate soil cleanup targets in interval and probabilistic cancer risk assessments. *Hum. Ecol. Risk Assess.*, 1, 89-100

Carrington, C.D. 1993. *Quantitative Risk and Response*, Contaminants Branch, Center for Food Safety and Applied Nutrition, US Food and Drug Administration, Washington, DC.

ChemRisk. 1994. Determination of Health-Based Soil CleanUp Levels using Probabilistic Analyses, Technical Report Prepared for Maxus Energy Corporation by Chem Risk Division of McLaren/Hart Environmental Engineering Corporation, Cleveland, OH, 10 October 1994

Clemen, R.T. 1991. *Making Hard Decisions*, Belmont, CA, Duxbury Press, Wadsworth Publishing Company.

Cooke, R.M. 1991. *Experts in Uncertainty, Opinion and Subjective Probability in Science*, Oxford, UK, Oxford University Press.

D'Agostino, R.B. and Stephens, M.A Eds *Goodness-Of-Fit Techniques*, New York, Marcel Dekker.

US Environmental Protection Agency. 1989 Human Health Evaluation Manual, Part A, Risk Assessment Guidance for Superfund, Volume I, Interim Final, EPA/540/1-89/002, Office of Emergency and Remedial Response, Washington, DC

Wolfram, S 1991 Mathematica™, A System for Doing Mathematics by Computer, Second Edition, Addison-Wesley, Redwood City, CA

## APPENDIX 1

### Feasible Regions in the $\mu$ - $\sigma$ Plane

In the fully probabilistic paradigm, the distribution for $\underline{\underline{R}}$ from Eqn 2 tends to a lognormal distribution as the number of input variables increases (regardless of the distributions of the input variables). In this appendix, we investigate how a risk management policy defined in terms of inequality constraints on various percentiles or moments of a lognormal distribution in turn places constraints on the two parameters of that lognormal distribution describing risk.

In this appendix, we use this notation for a lognormal distribution of risk: $\underline{\underline{R}} \sim \exp[\, N(\mu_R, \sigma_R)\,]$ where exp denotes the exponential function and $N(\bullet, \bullet)$ denotes a normal or Gaussian distribution with mean $\mu_R$ and standard deviation $\sigma_R$. For further information on this distribution, see Evans et al. (1993).

In the $\mu$ - $\sigma$ plane, we will investigate the constraints on $\mu$ and $\sigma$ that arise from mathematical principles and from different types of inequality constraints that risk managers may use to define acceptable risk. Each type of inequality constraint divides the $\mu$ - $\sigma$ plane into a feasible region and an infeasible region. The boundary between the two regions is the line of equality for the constraint. If multiple inequality constraints hold simultaneously (i.e., multiple inequality constraints are combined with the Boolean operator AND), the feasible region in the $\mu$ - $\sigma$ plane of the combination of constraints is the intersection of the feasible regions of the individual constraints.

First, we note one fundamental inequality constraint: $\sigma$ cannot be negative.

$$\sigma \geq 0$$

In all figures in this appendix, the feasible region for this constraint lies above the $\mu$-axis. When $\sigma = 0$, the random variable degenerates to a constant.

In Figure A-1, using Mathematica™ (Wolfram, 1991), we plot five illustrative inequality constraints on various percentiles of the distribution of risk. Here, we plot the straight lines for the five illustrative constraints for percentiles = 0.023, 0.159, 0.500, 0.841, and 0.977; these percentiles correspond to z = -2, -1, 0, +1, and +2, respectively.

Risk Analysis, Vol. 14, No. 4, 1994

# Principles of Good Practice for the Use of Monte Carlo Techniques in Human Health and Ecological Risk Assessments

David E. Burmaster[1] and Paul D. Anderson[2]

Received June 17, 1993; revised December 22, 1993

We propose 14 principles of good practice to assist people in performing and reviewing probabilistic or Monte Carlo risk assessments, especially in the context of the federal and state statutes concerning chemicals in the environment. Monte Carlo risk assessments for hazardous waste sites that follow these principles will be easier to understand, will explicitly distinguish assumptions from data, and will consider and quantify effects that could otherwise lead to misinterpretation of the results. The proposed principles are neither mutually exclusive nor collectively exhaustive. We think and hope that these principles will evolve as new ideas arise and come into practice.

KEY WORDS: Probabilistic risk assessment; Monte Carlo.

## 1. INTRODUCTION

For over 50 years, Monte Carlo (MC) techniques have been used in physics, chemistry, and many other disciplines to compute difficult multi-dimensional integrals. One example of this use is to combine probability distributions for several input variables to estimate probability distributions for one or more output distributions.[12,14] The widespread use of Monte Carlo techniques in public health and environmental risk assessment promises significant improvements in the scientific rigor of these assessments. Because Monte Carlo methods are more computationally intensive than the "deterministic" or "point estimate" methods in common use today, some people have suggested that Monte Carlo analysis not be widely adopted at this time. We believe that this is an overreaction, but we recognize the need for safeguards and precautions to reduce mistakes and prevent abuses.

[1] Alceon Corporation, P.O. Box 2669, Cambridge, Massachusetts 02238-2669.

[2] Ogden Environmental and Energy Services, 239 Littleton Road, Suite 7C, Westford, Massachusetts 01886.

We propose 14 principles of good practice in this article to assist people in performing and reviewing probabilistic risk assessments, especially in the context of the federal and state statutes concerning chemicals in the environment. Monte Carlo risk assessments for hazardous waste sites that follow these principles will be easier to understand, will explicitly distinguish assumptions from data, and will consider effects that could otherwise lead to misinterpretation of the results. These proposed principles arise from years of experience conducting and reviewing MC risk assessments and from conversations with many knowledgeable people in manufacturing companies, consulting companies, law firms, universities, nonprofit organizations, and government agencies. We think and hope that these principles will evolve as new ideas arise and come into practice.

Before proposing the 14 principles, we agree that each risk assessment, whether deterministic or probabilistic in design, must have a clearly defined assessment end point[9] and must contain all the information such that a knowledgeable person can reproduce and then evaluate the analysis from the material presented in the final report.[13]



0272-4332/94/0800-0477$07.00/1 

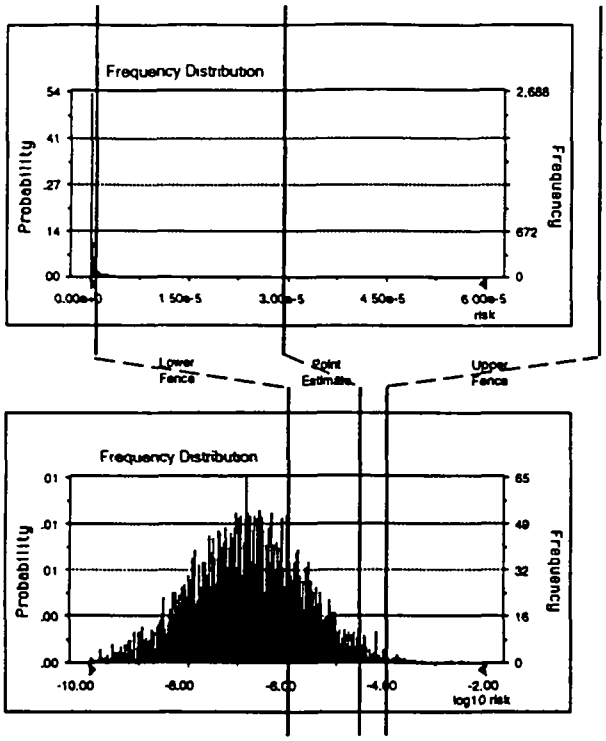



Fig. 1. Comparison of frequency distributions on linear and logarithmic scales.

ply missing information or to supplement partial information. If empirical measurements are not available for any reason, use and document accepted techniques—such as the Delphi method[3,13]—to estimate the input distributions for nonmeasured variables.

## 2.8. Principle 8

Discuss the methods and report the goodness-of-fit statistics for any parametric distributions for input variables that were fit quantitatively to measured data. Show plots of the parametric fits and the data on the same axes. Discuss the implications of any important differences. If any distribution was generated qualitatively or by expert judgment, discuss the techniques used.[18]

## 2.9. Principle 9

Discuss the presence or absence of moderate to strong correlations between or among the input variables. By strong correlation, we mean $|\rho| \geq 0.6$ or so. In many, but not all, practical situations, the absolute values of the correlations are less than 0.6. If so, the presence of moderate to strong correlations will have little effect on the central portions of output distributions[16] but may have larger effects on the tails of the output distributions. If it is possible that one or more moderate to strong correlations exist but no data are available from which to estimate them, perform Monte Carlo simulations with the correlations (i) set to zero and (ii) set to values considered high but plausible to learn if the possible correlations are important in the analysis. Display and discuss the results of these correlation sensitivity analyses and computational experiments, and state the practical effect, if any, of including or ignoring the correlations among the input variables.

## 2.10. Principle 10

Provide detailed information and graphs for each output distribution in the text and/or in an appendix. At a minimum, we suggest the following for each output variable: (i) a graph of the variable (in either log scale, linear scale, or both, depending upon the shape of the distribution) that clearly shows (a) the $10^{-4}$ risk and the $10^{-6}$ risk, or other allowable risk criteria, and (b) the point estimate of risk calculated by the deterministic method, and (ii) a table of the mean, the standard deviation, the minimum (if one exists), the 5th percentile, the median, the 95th percentile, and the maximum (if one exists). In Fig. 1, the histogram of estimated risk in the lower panel (on the log scale) gives a greater understanding of the variability in the output than does the histogram of the same results in the upper panel (on the linear scale). In Fig. 2, the histogram and the cumulative histogram in the upper and lower panels, respectively, display the variability of the output differently, but it is often useful to include both plots because each highlights a different aspect of the results. The graphs shown in Figs. 1 and 2 display the variabilities in the calculations, not the uncertainties.

## 2.11. Principle 11

Perform probabilistic sensitivity analyses for all of the key inputs represented by a distribution in the Monte Carlo analysis in such a way as to distinguish the effects of variability from the effects of uncertainty in the inputs. Display the results of these computational experiments in an appropriate graph.[9] The forms of the graphs will vary depending upon the method used to perform the probabilistic sensitivity analyses, but they should make clear which input variables contribute most

current conditions, a study to estimate risks for the reasonably foreseeable future conditions, and a study to estimate cleanup targets.

We have proposed these 14 principles of good practice as aids to performing or reviewing human health and ecological risk assessments done using MC techniques. While we favor the widespread use of MC techniques, we recognize the need for safeguards and precautions to reduce mistakes and prevent abuses. As proponents of the new methods, we hope that these proposed principles are general enough to show the standard of practice needed for conducting a MC assessment. We further hope that these ideas promote careful studies and innovation, which, in turn, create new insights and principles of good practice.

Several limitations apply to the ideas in this paper. First, the principles proposed are not mutually exclusive; some overlap with each other. Second, the principles proposed are not collectively exhaustive; for example, we have not proposed a principle concerning model uncertainty[13] nor one concerning the truncation of unbounded parametric input distributions (although the effects of truncation on percentiles and moments may be investigated through computational experiments and sensitivity analyses). Third, not all of these principles need apply to every study because not all of the principles are equally important in every situation. Fourth, the principles proposed are not inflexible recipes such as guidance manuals often present; we have instead tried to suggest the spirit of good practice without dictating a fixed and inviolate set of methods. Fifth, some of the principles are simply beyond the state of the art in some situations; for example, it is not now possible to fulfill all the proposed principles for a three-dimensional finite element model of time-varying ground water transport. Sixth, some of the principles are excessively burdensome for simple assessments. Notwithstanding all these limitations, we hope that the proposed principles will contribute to the quality of the MC studies undertaken. We further hope that these proposed principles will encourage others to refine these ideas to develop and publish new ones.

## ACKNOWLEDGMENTS

We thank Edmund A. C. Crouch, F. Owen Hoffman, Thomas E. McKone, Roy L. Smith, Alison C. Cullen, other colleagues, and two anonymous reviewers for helpful suggestions. Alceon Corporation and ENSR Consulting and Engineering supported this research.

## REFERENCES

1. K. T. Bogen. *Uncertainty in Environmental Risk Assessment* (Garland, New York, 1990).
2. D. E. Burmaster and K. von Stackelberg. "Using Monte Carlo Simulations in Public Health Risk Assessments: Estimating and Presenting Full Distributions of Risk," *J. Expos Anal. Environ. Epidemiol.* 1(4), 491–512 (1991).
3. N. C. Dalkey. *The Delphi Method: An Experimental Study of Group Opinion*, RM-5888-PR (Rand Corporation, Santa Monica, CA, June 1969).
4. A. M. Finkel. *Confronting Uncertainty in Risk Management, a Guide for Decision-Makers* (Center for Risk Management, Resources for the Future, Washington, DC, Jan. 1990).
5. H. C. Frey. *Quantitative Analysis of Uncertainty and Variability in Environmental Policy Making* (AAAS/US EPA Environmental Science and Engineering Fellows Program, American Association for the Advancement of Science, Washington, DC, 1992).
6. D. B. Hattis and D. E. Burmaster. "Some Thoughts on Choosing Distributions for Practical Risk Analyses" (submitted for publication).
7. B. Hayes. "The Wheel of Fortune, The Science of Computing," *Am. Sci.* **81**, 114–118 (1993).
8. F. O. Hoffman and J. S. Hammonds. *An Introductory Guide to Uncertainty Analysis in Environmental and Health Risk Assessment*, ESD Publication 3920 (Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, Oct. 1992).
9. F. O. Hoffman. "Propagation of Uncertainty in Risk Assessments: The Need to Distinguish Between Uncertainty Due to Lack of Knowledge and Uncertainty Due to Variability," U.S. EPA/University of Virginia Workshop on When and How Can You Specify a Probability Distribution When You Don't Know Much, University of Virginia, Charlottesville, VA, 19–21 Apr. (1993).
10. H. Ibrekk and M. G. Morgan. "Graphical Communication of Uncertain Quantities to Nontechnical People," *Risk Anal.* **7**, 519–529 (1983).
11. International Atomic Energy Agency. "Evaluating the Reliability of Predictions Using Environmental Transfer Models," Safety Practices Publications of the International Atomic Energy Agency, IAEA Safety Series, **100**, 1–106 (1989).
12. B. J. T. Morgan. *Elements of Simulation* (Chapman and Hall, London, 1984)
13. M. G. Morgan and M. Henrion. *Uncertainty, a Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis* (Cambridge University Press, New York, 1990).
14. R. Y. Rubinstein. *Simulation and the Monte Carlo Method* (John Wiley and Sons, New York, 1981).
15. A. Shlyakhter and D. M. Kammen. "Sea-Level Rise or Fall," *Nature* **357**, 25–7 (1992).
16. A. E. Smith, P. B. Ryan, and J. S. Evans. "The Effect of Neglecting Correlations When Propagating Uncertainty and Estimating Population Distribution of Risk," *Risk Anal.* **12**, 467–474 (1992).
17. G. W. Suter II. *Ecological Risk Assessment* (Lewis, Chelsea, MI, 1993)
18. A. C. Taylor. "Using Objective and Subjective Information to Develop Distributions for Probabilistic Exposure Assessment," *J. Expos. Anal. Environ. Epidemiol.* 3(3), 285–298 (1993).

# USING MONTE CARLO SIMULATIONS IN PUBLIC HEALTH RISK ASSESSMENTS: ESTIMATING AND PRESENTING FULL DISTRIBUTIONS OF RISK

**DAVID E. BURMASTER, PH.D.**
**KATHERINE VON STACKELBERG**
**Alceon Corporation**
**Cambridge, Massachusetts**

*With desktop computers as powerful as mainframes were just a few years ago, analysts can now use commercial software to estimate full probability distributions for—not just point estimates of—health risks experienced by people chronically exposed to toxic chemicals at or near hazardous waste sites Even though probability is the central concept in risk assessment, and even though probabilistic methods offer strong advantages and insights as compared to the "deterministic" methods now required by U.S. Environmental Protection Agency's guidance manuals, analysts have only begun to use probabilistic methods at Superfund sites*

*In this paper, we examine a simplified case study using Monte Carlo methods to estimate full distributions of public health risk We demonstrate the use of "toggles" to isolate the contributions of different inputs, and we also offer new graphical methods to communicate the results to risk managers and concerned citizens.*

## INTRODUCTION

Risk assessments that follow guidance published by the U.S. Environmental Protection Agency (EPA) combine a series of average, conservative, upperbound, and worst-case assumptions to derive a point estimate of risk that is conservative, i.e., protective of public health (EPA, 1989a; EPA, 1989b). Although EPA calls for analyses which address Reasonable Maximum Exposure (RME) to receptors, the concept of Reasonable Maximum Exposure is never fully defined (EPA, 1989a;b).

Conservative point estimates of risk calculated with EPA's current methods have three major limitations. First, by selecting a combination of average, conservative, and worst-case

1. Direct all correspondence to David E. Burmaster, Ph.D., Alceon Corporation, P.O. Box 2669, Harvard Square Station, Cambridge, MA 02238-2669. Tel. (617) 864-4300

Journal of Exposure Analysis and Environmental Epidemiology, Vol. 1, No 4, pp 491-512

ISSN: 1053-4245

assumptions, risk assessors and risk managers have no way of knowing the degree of conservatism in an assessment. Second, by setting the *bias* high enough to swamp the uncertainty for each of many variables, the risk assessment considers scenarios that will rarely if ever happen. Third, it is meaningless to run traditional sensitivity analyses (e.g., to make calculations at ±10 or ±25 percent from each input value) to determine the uncertainties in the final point estimates because many of the input variables are at or near their maxima. Thus, the current procedures offer comfort if the estimated risks falls below a *de minimis* value, but they offer no insight if the estimated risk exceeds a *de minimis* value.

Because conservative assumptions usually combine in multiplicative ways, results based on EPA's current methods rarely if ever capture the true risk of a situation or a behavior. The calculated point estimate usually falls far above the $95^{th}$ percentile of the true risk range (Environ, 1991). Monte Carlo simulations can estimate the full risk distributions, thereby putting the point estimates into a full and proper context.

Monte Carlo simulation, developed by physicists over 50 years ago and long used by engineers in many fields, addresses the weaknesses of the current risk assessment methods identified above (Burmaster et al., 1990). In extending the regular methods used in public health and ecological risk assessments, probabilistic techniques add several steps to estimate both point values and full distributions for the exposures and risks (Smith, 1991). First, the analyst determines a continuous or discrete probability density function (PDF) to describe each of the variables to be included in the analysis. In this step, the analyst must also determine if any correlations exist among the input variables and take appropriate action if necessary. Second, the analyst uses suitable software to make many realizations of the model. For each realization, the computer draws one random value from the appropriate distribution for each of the random variables in the model, and computes and stores a single result. This computation is repeated many times. Third, the analyst views the results and establishes the shapes of the distributions for intermediate and final results, and various statistical summaries of the results. In this framework, a complete risk distribution is derived by combining the distributions for the antecedent variables. These probabilistic techniques make the analyses more informative for risk managers and members of the public (Finkel, 1990). These new methods are illustrated in the first two fully probabilistic risk assessments prepared for hazardous waste sites regulated under the federal Superfund program (Ebasco, 1990; Environ, 1991).

In Monte Carlo simulation, each of many input variables can become a random variable with known or estimated PDF. (Equivalently, an input variable can be specified by a cumulative distribution function (CDF)). Within this framework, a variable takes on a range of values with a known probability.

The histograms for the estimated risks from exposure to a single compound are often highly non-Gaussian in shape for two reasons. First, some or all of the input variables may not have

normal or even symmetric distributions. Second, the input variables usually enter the formulae by multiplication and division (and subsequent summation), so that even if all inputs have Gaussian distributions, the results will not. For risk from exposure to a single compound, the central limit theorem of statistics implies that the product of many distributions tends to a lognormal distribution, regardless of the distributions of the individual factors (see a statistical text, e.g., Benjamin and Cornell, 1970).

## A SIMPLIFIED CASE STUDY

To illustrate the application of Monte Carlo simulation to health risk assessment, we consider a simplified case study of a hypothetical site. We estimate the PDFs and summary statistics for the Incremental Lifetime Cancer Risk (ILCR) for one scenario involving dermal exposures to benzo(a)pyrene (BaP) found in soils. We chose point values and distributions for the inputs that are reasonable in view of the current knowledge and current EPA guidance documents.

As a case study, we consider a hypothetical site — an old industrial site with BaP in the surface soils — that a City Council may buy and convert to a park. Since we want to illustrate the use of Monte Carlo simulation to estimate a full distribution for health risk, we consider only one of the many scenarios and only one of many possible exposure pathways which could be considered for this site and its proposed use. This scenario considers children who are exposed to surface soils while playing in the new park. We make conservative and simplifying assumptions concerning the children's dermal contact with the soil. Given the uncertainties inherent in an exposure assessment, this scenario is constructed in accordance with current EPA guidelines and uses conservative (or health-protective) assumptions, in the spirit of analyzing the RME case, rather than worst-case assumptions.

## THE EXPOSURE MODEL

To estimate health risks, we first estimate the average daily dose of BaP that a person receives in units of milligram of bioavailable chemical per kilogram of body weight per day (mg/(kg•d)), averaged over a 70-year life (abbreviated as the ADD(life)). Following the standard method (EPA, 1989b), we then estimate the Incremental Lifetime Cancer Risk by multiplying the ADD(life) by the Cancer Potency Factor (CPF). Exhibit 1 gives the equations used to estimate ADD(life) and ILCR for this case study.

McKone recently published a model to estimate the uptake of organic chemicals from a soil matrix deposited onto the skin surface (McKone, 1990). In the model, the stratum corneum is the barrier to uptake, and the amount of chemical which passes through the stratum corneum represents the bioavailable dose. The model depends on scenario-specific inputs, soil properties, skin properties, and chemical properties of the soil contaminants. We use the one-time or unit-deposition model in this simplified analysis. McKone derives a Personal Exposure Factor (PEF) which, when multiplied by the concentration of the chemical in the

soil, estimates the average daily dose on a day of exposure. We identified the most sensitive variables in McKone's model by standard sensitivity analyses.

## THE SPREADSHEET MODEL

Exhibit 2 shows the spreadsheet for estimating doses and risks for this case study. Left of the vertical bar, the spreadsheet lists all the variables considered in this analysis, along with point estimates for the variables and parameters for the probability distributions. For clarity and subsequent analysis, we group the input variables as shown: exposure scenarios, soil properties, skin properties, chemical properties, soil concentrations, relative bioavailability, and cancer potency factor. For each of the 12 variables with probability inputs, the spreadsheet has a 0,1 toggle to select between the point value (activated by 0) and the PDF (activated by 1).

Right of the vertical bar, the spreadsheet calculates intermediate results and reports the estimated ILCR ( and $\log_{10}$ ILCR) in the lower right corner. As shown here, all toggles are set to 0 and thus the spreadsheet has calculated the point estimate for ILCR. The spreadsheet estimates the Incremental Lifetime Cancer Risk from chronic low-dose exposure to BaP via dermal contact with soils, in keeping with the methods recommended by EPA (EPA, 1989a;b). In the absence of specific information on possible synergisms or antagonisms among carcinogenic compounds, the total ILCR is estimated by summing the values for each compound over all pathways (although only one pathway is specified for this analysis).

## INPUT VALUES AND DISTRIBUTIONS

With the exposure model complete, we identify point estimates for all of the model inputs, find in the literature or formulate distributions for the inputs we want to vary, and put all of the information into an appropriate simulation program. For use in the exposure model, we formulate distributions for the concentration of BaP in the site soils and the CPF. Based on the results of the sensitivity analysis, we formulate distributions for key variables in the McKone model: body weight, the time soil stays on skin, average body surface area, fraction of skin area exposed, soil loading, bulk density of soil, and skin water content. We also formulate distributions for exposure days per week, exposure weeks per year, exposure years per life, BaP soil concentration, and the CPF for BaP. Thus, five of the seven groups of variables—exposure scenario, soil properties, skin properties, soil concentrations, and cancer potency factor—include one or more variables which a toggle can switch between a point estimate and a distribution.

In this case study, we use three common distributions to describe the key model inputs: the normal or Gaussian distribution, the lognormal distribution, and the uniform distribution. We denote random variable X with a normal distribution as X ~ Normal ($\mu$, $\sigma$), where $\mu$ and $\sigma$ represent the arithmetic mean and standard deviation, respectively. Similarly, the lognormal distribution is denoted as X ~ Lognormal (m, s) where m and s represent the arithmetic mean

and standard deviation of the underlying normal distribution, respectively. (The underlying normal distribution is generated by taking the logarithms of the values in the distribution). Finally, we use the notation X ~ Uniform (a, b) to show that the random variable X is distributed uniformly between fixed minimum (a) and maximum (b) values.

Exhibit 2 shows all the inputs chosen for the point values and the distributions, along with a reference. All of the point values are reasonable in the sense that EPA has or could readily endorse the values for a particular site.

Most people intuitively understand that some or all of the variables in the various groups are truly stochastic in nature. We go further than most analysts though, and we consider that CPF values are also stochastic. After all, EPA discusses CPF values in probabilistic language as representing the 95$^{th}$ percentile of slope of the linearized multistage model applied to animal data and extrapolated to humans. Extending the ideas in earlier publications (Crouch, 1983; Crouch & Wilson, 1981), Crouch re-evaluated the CPF for BaP (Crouch, 1990). Based on this information, we model the ingestion CPF for BaP with a lognormal distribution: $CPF_{BaP}$ ~ Lognormal (-0.79, 2.39) in units of $(mg/(kg \cdot d))^{-1}$). Although EPA has never published an ingestion CPF for BaP in its Integrated Risk Information System, the value it now uses in practice, 11.5 $(mg/(kg \cdot d))^{-1}$) (EPA, 1986), falls at approximately the 91$^{st}$ percentile of Crouch's distribution. (We note that the cross-assignment of the CPF from the ingestion pathway to the dermal contact pathway, though accepted in practice, is incorrect.)

## ESTIMATION AND PRESENTATION OF RISKS

We now estimate full distributions of health risks for this case study, using commercial software (Crystal Ball, V2.0 (Decisioneering, 1991)) in conjunction with the spreadsheet, and we compare the distributions to the point estimate of risk. We consider different ways to present the risks in a graphic format to both non-technical and technical audiences, as drawn from various widely recognized sources (Chambers et al., 1983; Cleveland, 1985; Finkel, 1990; Graham and Henrion, 1984; Ibrekk and Morgan, 1987; Tufte, 1983; 1990; Tukey, 1977; Systat, 1991).

*The Deterministic Case*

The ILCR shown in Exhibit 2, namely 2.96E-05, is the "conservative point estimate" calculated by combining the point values of all the inputs. Although this point estimate of risk is the usual stopping point for risk assessments, we compare it to the full distributions. We also compare EPA's target risk range to the full distributions, i.e., the range from: ILCR = $10^{-4}$, the risk at which EPA always requires remediation, to ILCR = $10^{-6}$, the Agency's "point of departure" for remedial goals under the Superfund Program (EPA, 1990). In Exhibits 3-13, the three vertical lines locate the limits of EPA's target risk range and the point estimate on the full distributions.

*The Probabilistic Case with All Distributions Activated*

Exhibits 3-7 show different ways to present the results of 5,000 realizations of the spreadsheet model with distributions for all 12 random variables activated. Each method highlights a different aspect of the results, and each has different strengths and weaknesses, depending on the technical sophistication of the audience.

Exhibit 3 summarizes the full distribution of risk as a table of statistics, including the mean, median, mode, standard deviation, and deciles for the ILCR and the $\log_{10}$ ILCR. As a nongraphical method, this approach has only slightly more appeal than an ordered list of the 5,000 results! Even engineers have difficulty interpreting the results when presented in this fashion, although it is possible to discern that the conservative point estimate falls above the $95^{th}$ percentile of the full risk distribution.

Exhibit 4 compares the histograms of the ILCR and $\log_{10}$ ILCR for the case study. The graphs reveal important features not evident in the previous table. In linear space, the full distribution has a long right tail, a high variance, and a mode far below the conservative point estimate. In this upper histogram, only 4,941 of the results from the 5,000 realizations are visible within the domain in the graph. In logarithmic space, the full distribution has more symmetry, with the qualitative feel of a normal distribution. In logarithmic space, it is easier to grasp the relationships among the distribution, the conservative point estimate, and EPA's target risk range of $10^{-4}$ to $10^{-6}$. In this lower histogram, more realizations (4,990 of 5,000) are visible in the graph, but some still fall outside the domain graphed. Each histogram in Exhibit 4 confirms that the conservative point estimate falls well above the $95^{th}$ percentile of the full risk distribution.

Exhibit 5 compares the ordinary histogram and the cumulative histogram for the $\log_{10}$ ILCR for this case study. Although these graphs contain identical information, an informal poll revealed that non-technical audiences understand the ordinary histogram far more readily than the cumulative histogram and that technical audiences prefer to have both presented. Again, the vertical lines locate the conservative point estimate and EPA's target risk range on the full distribution, and again, only 4,990 results from the 5,000 realizations are visible.

Exhibits 6 and 7 show a "box-and-whiskers" diagram and a probability plot of the $\log_{10}$ ILCR, complete with lines to locate the conservative point estimate and EPA's target risk range on the full distribution. As presented here, the box-and-whiskers plot marks the minimum and maximum at the ends of the whiskers, the $10^{th}$ and $90^{th}$ percentiles as the short crossbars, the $25^{th}$ and the $75^{th}$ percentile as the ends of the box, and the median as the crossbar near the center of the box. For a technical audience, the probability plot in Exhibit 6 shows that the 5,000 realizations of $\log_{10}$ ILCR closely follow a lognormal distribution for 2.5 or 3 standard deviations above and below the median.

*The Probabilistic Case Showing the Contributions from Different Inputs*

Uncertainties propagate and combine through a series of calculations. While it is highly unlikely that the variabilities will combine in a purely additive or multiplicative way to produce the theoretically largest possible uncertainty, it is true that the overall uncertainty in a calculation can never be smaller than the uncertainty associated with the least certain step in the chain. For this reason, it is useful to disaggregate the contributions from the different groups of input variables, using the toggles to isolate the effects.

Exhibit 8 shows the contributions of the variables in each of five input groups in a table of statistics similar to those for $\log_{10}$ ILCR in Exhibit 3. In other words, Exhibit 8, a novel type of probabilistic sensitivity analysis, tabulates the distributions of $\log_{10}$ ILCR that result from random realizations of each of the five input groups while keeping the other four constant. Again, this table has little or no intuitive appeal, although it does demonstrate that the conservative point estimate combines assumptions that exceed the $95^{th}$ percentile for the variables in the exposure group and exceed the $90^{th}$ percentile for the CPF. To visualize the results of this probabilistic sensitivity analysis in graphs, we present several different views of the same information in Exhibits 9-13, each with lines to locate the conservative point estimate and EPA's target risk range for comparison.

Exhibits 9 and 10 parallel earlier exhibits but also show the contributions from each group of variables. In each exhibit, the top histogram shows the distribution with all the toggles activated as a frame of reference. The second panel shows the risk distribution with only the toggles in the exposure group activated. The remaining panels in Exhibits 9 and 10 show the contributions from the variables in the remaining groups, as labeled. Exhibit 11 condenses the cumulative distributions from the previous exhibit. Exhibits 12 and 13 show the box-and-whiskers plots and the probability plots for each contributing group of input variables.

In different ways, Exhibits 9-13 portray the same information, and each allows us to understand the contributions of each of the groups of input variables in different ways. In declining order, the greatest variabilities and uncertainties flow from the CPF variable, the variables in exposure group, and the concentration variable. We see support for this assertion in the relative widths of the histograms in Exhibit 9, the relative steepness of the cumulative distribution in Exhibits 10 and 11, the relative widths of the boxes and the whiskers in Exhibit 12, and the relative slopes of the probability plots in Exhibit 13. Interestingly, these same graphs show that EPA makes its most conservative policy assumptions to compensate for the uncertainty of the variables with the greatest contributions.

## DISCUSSION

We promote the use of spreadsheets and Monte Carlo software in practical applications, even though some simple situations (e.g., the multiplication of many variables distributed lognormally) can be calculated or well approximated by closed-form expressions (Shlaykhter,

1991, citing Broadbent, 1956, and Fenton, 1960) for two reasons. First, most practical risk assessments, such as those for hazardous waste sites, are far more complicated than can be addressed analytically. Second, few risk assessors have the mathematical skill to manipulate the closed-form methods (see Springer, 1979).

New software used in conjunction with spreadsheets on powerful desktop computers provides an easy and fast way to estimate probability distributions for human health risks in the assessments of sites with chemical contamination. While the methods are straightforward and can easily be extended to linked spreadsheets and correlated input variables (Decisioneering, 1991), continued research is needed to specify input distributions for exposure-related variables and new methods are needed to quantify the distributions appropriate for CPFs.

In this simplified case study, we note that the point estimate is truly conservative because calculating the point estimate compounds many conservative assumptions and values. When the distribution of ILCR is presented as a histogram with the point estimate demarked, the distribution is barely visible in linear space. For this reason, we recommend that risk assessors also present the estimated distributions with point estimates demarked in logarithmic space in a way that reveals the order of magnitude of the results.

Although we have chosen a simplified case study to illustrate the calculation and presentation of distributions of exposure and risk, commercial software can be applied to far more general cases, including: multiple compounds, multiple exposed populations, multiple exposure pathways, and multiply correlated exposure variables (through the technique of Iman and Conover (1982) and Iman and Davenport (1982)). With the combination of Excel™ and Crystal Ball™, an analyst can design and perform a probabilistic risk assessment using any set of algebraic, equilibrium, or steady state models in linked spreadsheets. As a general proposition, any situation that can be modeled in Excel™, a full-featured spreadsheet, can then be simulated in Crystal Ball™ by specifying either common distributions from built-in "Gallery" (e.g., uniform, triangular, normal, lognormal, exponential, weibull, gamma, beta, poisson, binomial, geometric, and hypergeometric) or a custom distribution (using (x,y) pairs to specify the breakpoints of a piecewise linear CDF). By combining the features of the spreadsheet and the simulation software, an analyst can also model mixture problems and can derive new distributions, e.g., the distribution of a function, say, $f(X) = X^{2.3}$, where X is distributed lognormally.

## RECOMMENDATIONS

Reviewing the various ways to present the overall distribution for ILCR, we recommend a combination of the ordinary histogram and the cumulative histogram as shown in Exhibit 5 (see similar recommendation in Ibrekk and Morgan, 1987), perhaps in combination with a box-and-whiskers plot to the same scale along the lower edge of the ordinary histogram. We find some highly technical audiences prefer the graph in Exhibit 7. We also recommend that

analysts include a variety of toggles in each model to isolate the contribution from each of several groups of stochastic variables. To display these results, we recommend the graphs in Exhibits 9 and 10 (or 11), perhaps supplemented with Exhibit 13.

We urge EPA to endorse and encourage the use of Monte Carlo simulations as a way to supplement and eventually replace current methods. We believe Monte Carlo analysis separates risk assessment from risk management in the sense originally recommended by the National Academy of Sciences (1983). Monte Carlo techniques provide a method to estimate the distribution of risk and to understand the degree of conservatism present in a point estimate.

## LIMITATIONS

While we believe the strengths of the Monte Carlo methods far outweigh any limitations, this case study rests on many assumptions which simplify the calculations but which also limit the results. While it is not possible to list all the simplifications, it is important to discuss some of the main types and to give illustrations.

- First, this case study uses greatly simplified equations to estimate exposure to chemicals. Although the equations follow current federal guidance for public health risk assessments, they are dramatic simplifications of reality. A risk distribution based on a model is only as good as the model. In this simplified case study, for example, we use only one model for dermal exposures and we assign the CPF for ingestion exposures to dermal exposures. Each of these two assumptions is the source of additional uncertainty quite outside the model and the results. The analyst must acknowledge exogenous sources of uncertainty (not included in the model) and discuss their potential for shifting and/or increasing the variance of the estimated risk distribution.

- Second, this case study ignores obvious correlations among variables. As an example, body weight and skin area are certainly correlated, and the joint distribution of these variables is undoubtedly a function of age and sex .

- Third, even in the Monte Carlo simulations, the case study treats many variables known to be stochastic as deterministic. While it is relatively easy to overcome the third class of oversimplification and limitation within current knowledge and computational resources, more research is needed to address and resolve the limitations imposed by the first two classes of simplifying assumptions.

- Finally, although it may seem obvious, inadvertent or deliberate abuse of the Monte Carlo approach can occur and does lead to unrealistic and incorrect results. With a powerful new tool available for use, we must all strive to use it wisely and appropriately, especially in regard to the specification of input distributions and functional relationships.

## ACKNOWLEDGMENTS

We thank Nancy Irwin Maxwell for many helpful comments and suggestions. We also thank two anonymous reviewers for their careful readings and excellent suggestions. Alceon Corporation funded this research.

## TRADE MARKS

Crystal Ball™ is a registered trademark of Decisioneering, Inc.

## EXHIBITS

### EXHIBIT 1.

### Equation to Model Dermal Exposure to Soils

$$ADD(life) = \frac{Cs \cdot PEF \cdot DpW \cdot WpY \cdot YpL}{DinY \cdot YinL}$$

where:

| | | |
|---|---|---|
| ADD(life) | = | Average Daily (bioavailable) Dose, averaged over a lifetime (mg/(kg•d)), |
| Cs | = | soil exposure concentration (mg/kg), |
| PEF | = | personal exposure factor, averaged over a day of exposure (kg/(kg•d)), |
| DpW | = | exposure days per week (d/wk), |
| WpY | = | exposure weeks per year (wk/yr), |
| YpL | = | exposure years per lifetime (yr/life), |
| DinY | = | total number of days per year [7 (d/wk) • 52 (wk/yr)], and |
| YinL | = | total number of years per lifetime (70 yr/life) |

### Equation to Estimate Incremental Lifetime Cancer Risk

$$ILCR = ADD(life) \cdot CPF$$

where:

| | | |
|---|---|---|
| ILCR | = | Incremental Lifetime Cancer Risk, the incremental probability that a person will develop cancer during lifetime (probability) |
| ADD(life) | = | Average Daily Dose of a compound, averaged over life during which exposure occurs, (mg/(kg•d)) |
| CPF | = | Cancer Potency Factor for a compound, by ingestion ($(mg/(kg \cdot d))^{-1}$) |

## EXHIBIT 2. Spreadsheet for Dermal Contact (One-Time Deposition)

| Variables | Units | Point Estimate | Distribution | Parameters | Parameter Values first | second |
|---|---|---|---|---|---|---|
| Exposure Scenario: | | | | | | |
| average body weight | kg | 47 | Normal | (μ,σ) | 47 | 8 3 |
| time soil stays on skin | hr | 6 | Normal | (μ,σ) | 6 | 1 |
| average body surface area | m2 | 1 4 | Normal | (μ,σ) | 1 4 | 0 17 |
| fraction of skin area exposed | frac | 0.2 | Lognormal | (m,s) | -2.15 | 0 50 |
| skin soil loading | mg/cm2 | 1 | Uniform | (a,b) | 0 75 | 1 25 |
| exposure days per week | d/wk | 1 | Uniform | (a,b) | 0 5 | 1 5 |
| exposure weeks per year | wk/yr | 20 | Uniform | (a,b) | 5 | 25 |
| exposure years per life | yr/ life | 10 | Uniform | (a,b) | 5 | 12 |
| Soil Properties | | | | | | |
| soil bulk density, Rho(b) | kg/m3 | 1600 | Normal | (μ,σ) | 1600 | 80 |
| soil porosity, Phi | m3/m3 | 0 5 | | | | |
| soil water content, Theta | m3/m3 | 0 3 | | | | |
| organic carbon fraction, foc | frac | 0 02 | | | | |
| Skin Properties | | | | | | |
| skin thickness, Delta(skin) | m | 1 5E-05 | | | | |
| skin fat content, f(fat) | kg/kg | 0 1 | | | | |
| skin water content, Gamma | m3/m3 | 0 3 | Normal | (μ,σ) | 0 3 | 0 05 |
| boundary layer size, Delta(a) | m | 0 0045 | | | | |
| Chemical Properties | | | | | | |
| benzo(a)pyrene Kow | frac | 1 55E+06 | | | | |
| benzo(a)pyrene Kh | frac | 2.04E-05 | | | | |
| D(air) | m2/s | 5E-06 | | | | |
| D(water) | m2/s | 5E-10 | | | | |
| Soil Concentrations* | | | | | | |
| benzo(a)pyrene | mg/kg | 29 49 | Lognormal | (m,s) | 2 81 | 0 68 |
| Relative BioAvailability | | | | | | |
| benzo(a)pyrene | frac | 0 3 | | | | |
| Cancer Potency Factor | | | | | | |
| benzo(a)pyrene | (kg•d)/mg | 11 5 | Lognormal | (m,s) | -0 79 | 2 39 |

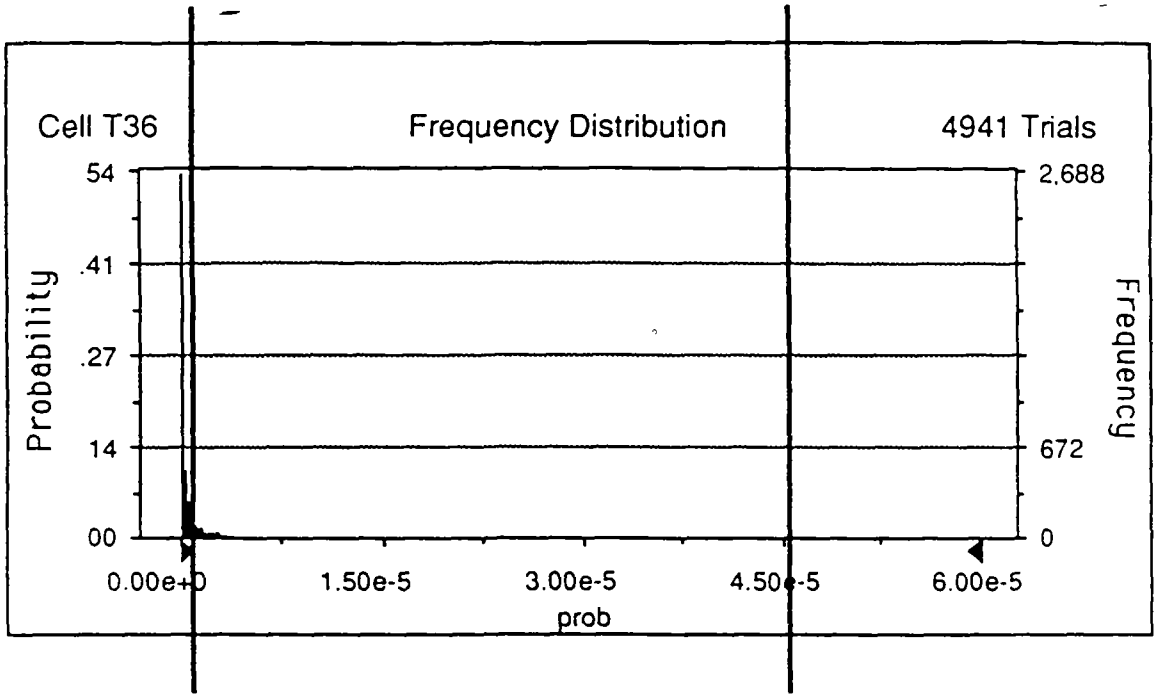
## EXHIBIT 2. Spreadsheet for Dermal Contact (One-Time Deposition) (cont'd)

| Toggle (0,1) | Reference | Variables | Units | Results |
|---|---|---|---|---|
| | | General: | | |
| 0 | Team, 1991 (GCA, 1984) | skin area exposed | m2 | 0 28 |
| 0 | Team, 1991 | soil air content, Alpha | m3/m3 | 0 2 |
| 0 | Team, 1991 (GCA, 1984) | soil total density, Rho(t) | kg/m3 | 1900 2 |
| 0 | Team, 1991 (GCA, 1984) | D(G) | m2/s | 9 41E-08 |
| 0 | Team, 1991 | D(L) | m2/s | 3.63E-11 |
| 0 | Team, 1991 | soil deposition | mg | 2800 |
| 0 | Team, 1991 | soil dep | kg | 0 0028 |
| 0 | Team, 1991 | del(soil) | m | 5 26E-06 |
| | | Dermal Contact | | |
| 0 | Team, 1991 | RG(soil) | | 1 17E+09 |
| | McKone,1990 | RL(soil) | | 23,808 30 |
| | McKone,1990 | D(soil), DG/RG + DL/RL | m2/s | 1 61E-15 |
| | McKone,1990 | D(skin) | m2/s | 6 48E-16 |
| | | K(vapor loss from skin) | | 9 52E-13 |
| | | Kd | | 14,880 00 |
| | | Uptake fraction | | 0 19 |
| | | Ku or Keff(sl-sk) | | 3 78E-11 |
| | McKone,1990 | c | | 0 03 |
| | McKone,1990 | b | | 0 03 |
| 0 | Team, 1991 | PEF (d) | kg/(kg•d) | 1 11E-05 |
| | McKone,1990 | PEF (life) | kg/(kg•d) | 8 72E-08 |
| | | ADD(life) | mg/(kg•d) | 2.57E-06 |
| | | BaP dermal contact ILCR | prob | 2 96E-05 |
| | | BaP dermal contact ILCR | log10 | -4 53 |
| | Mabey, 1982 | | | |
| | Mabey, 1982 | | | |
| | McKone,1990 | | | |
| | McKone,1990 | | | |
| 0 | Team, 1991 | | | |
| | US EPA, 1989, I | | | |
| 0 | US EPA, 1986, SPHEM, Crouch,1990 | | | |

**EXHIBIT 3. Summary Statistics with All Toggles On**

| | In Linear Space | In Log 10 Space |
|---|---|---|
| | ••••• | ••••• |
| Trials | 5,000 | 5,000 |
| Mean | 1.78E-06 | -6.80 |
| Median | 1.84E-07 | -6.81 |
| Mode | 1.00E-07 | -6.84 |
| Standard Deviation | 5.52E-06 | 1.13 |
| Variance | 3.05E-11 | 1 28 |
| Skewness | 5 62 | 0 03 |
| Kurtosis | 40 26 | 2.85 |
| Range Width | 3 35E-03 | 9 84 |
| Minimum | 4 88E-13 | -12.31 |
| Maximum | 3 35E-03 | -2.47 |
| Mean Standard Error | 0 | 0.02 |
| | | |
| Percentile | | |
| 0 | 4 88E-13 | -12 31 |
| 5 | 5.89E-07 | -8 70 |
| 10 | 1 18E-06 | -8 27 |
| 15 | 1 77E-06 | -8 00 |
| 20 | 2 35E-06 | -7 79 |
| 25 | 2.94E-06 | -7 58 |
| 30 | 3 53E-06 | -7 40 |
| 35 | 4 12E-06 | -7 25 |
| 40 | 4 71E-06 | -7 09 |
| 45 | 5 30E-06 | -6.95 |
| 50 | 5 89E-06 | -6 81 |
| 55 | 6 48E-06 | -6.67 |
| 60 | 7 06E-06 | -6 53 |
| 65 | 7 65E-06 | -6 37 |
| 70 | 8 24E-06 | -6 21 |
| 75 | 8 83E-06 | -6 04 |
| 80 | 9 24E-06 | -5 83 |
| 85 | 1 00E-05 | -5 61 |
| 90 | 1 06E-05 | -5 35 |
| 95 | 1 14E-05 | -4 96 |
| 100 | 3 35E-03 | -2 47 |

**EXHIBIT 4. Comparison of Frequency Distributions on Linear and Logarithmic Scales**



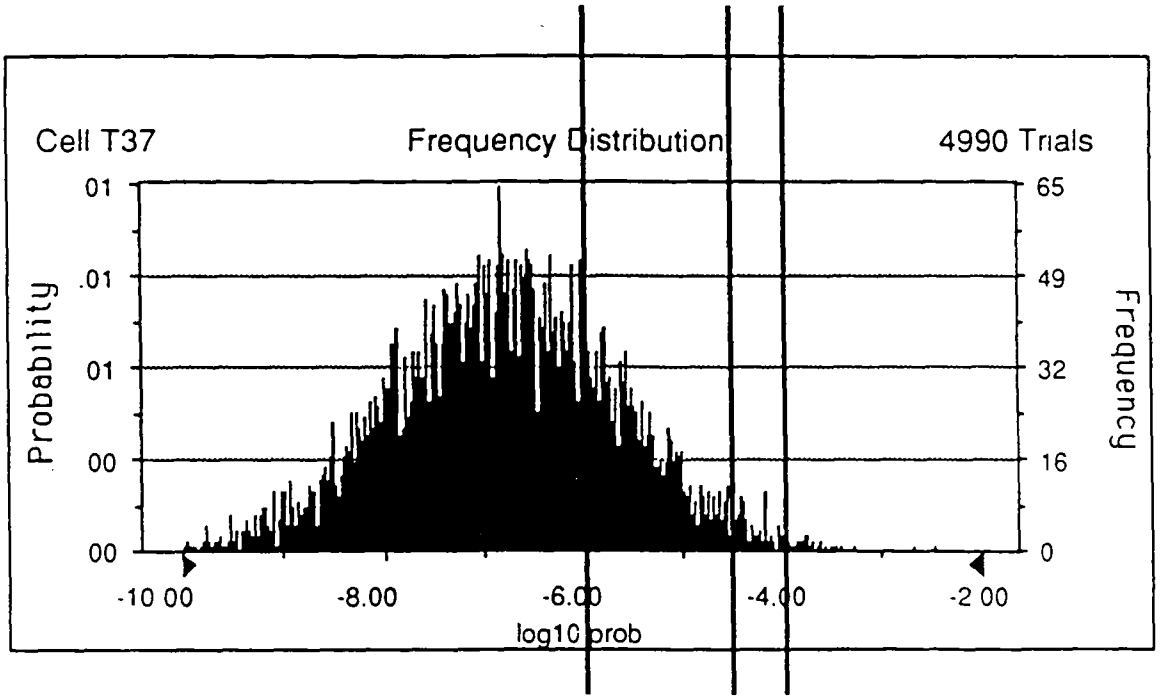



**EXHIBIT 5. Comparison of Frequency Distribution and Cumulative Distribution on Logarithmic Scale**

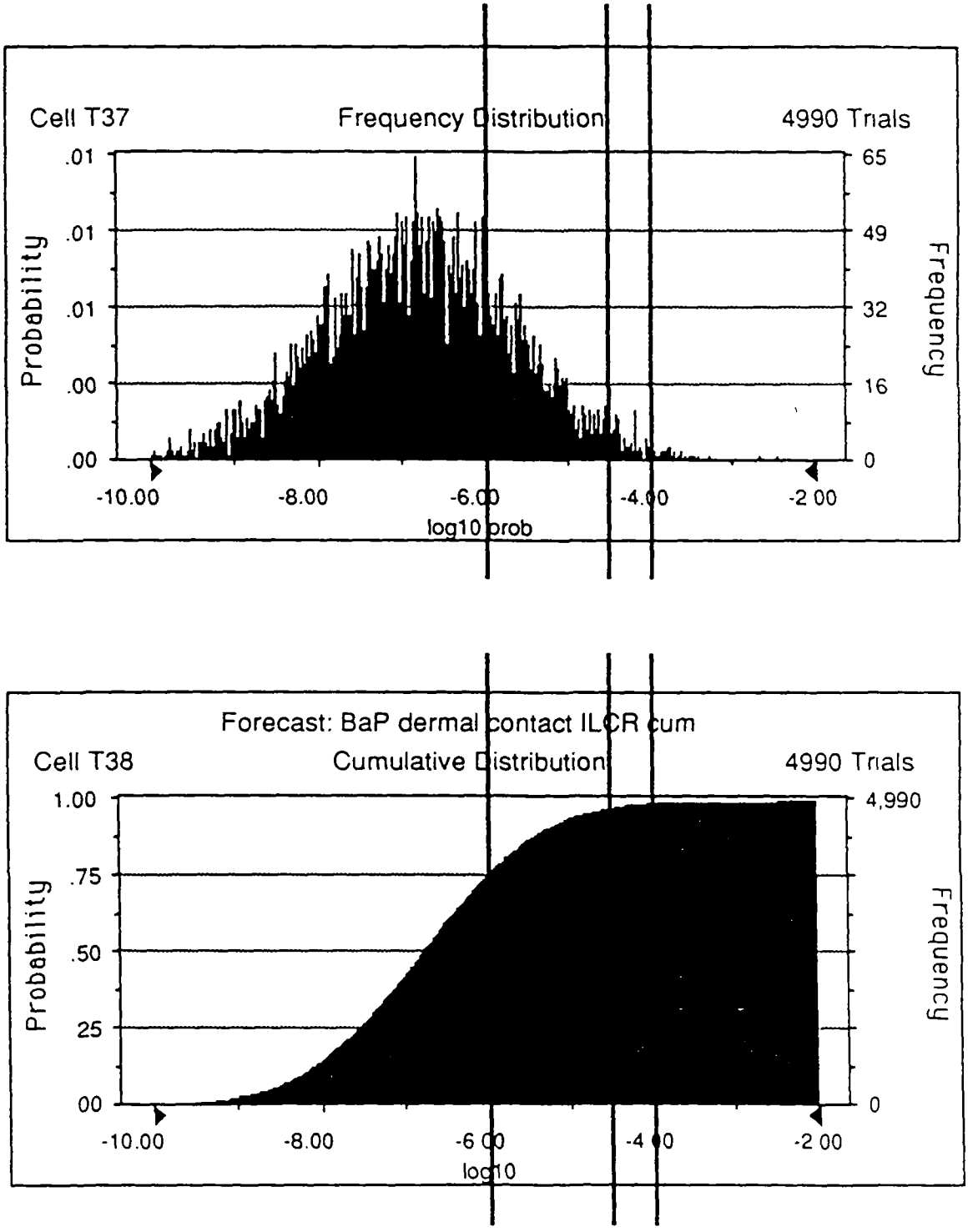


**EXHIBIT 6. Box and Whiskers Diagram with All Toggles On**

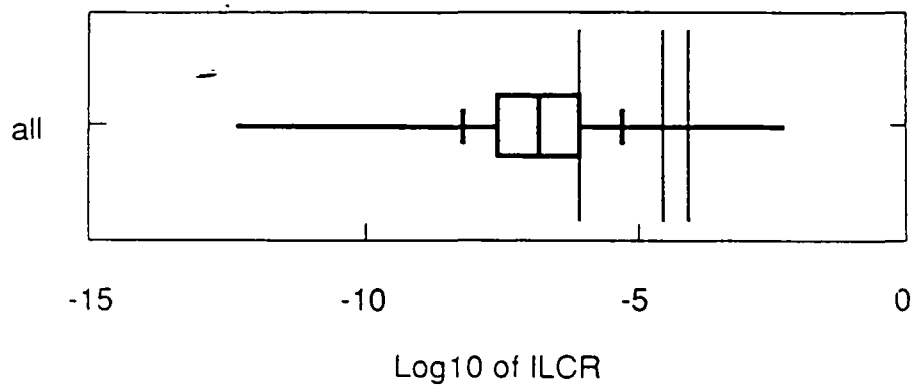



**EXHIBIT 7. Probability Plot of ILCR with All Toggles On**

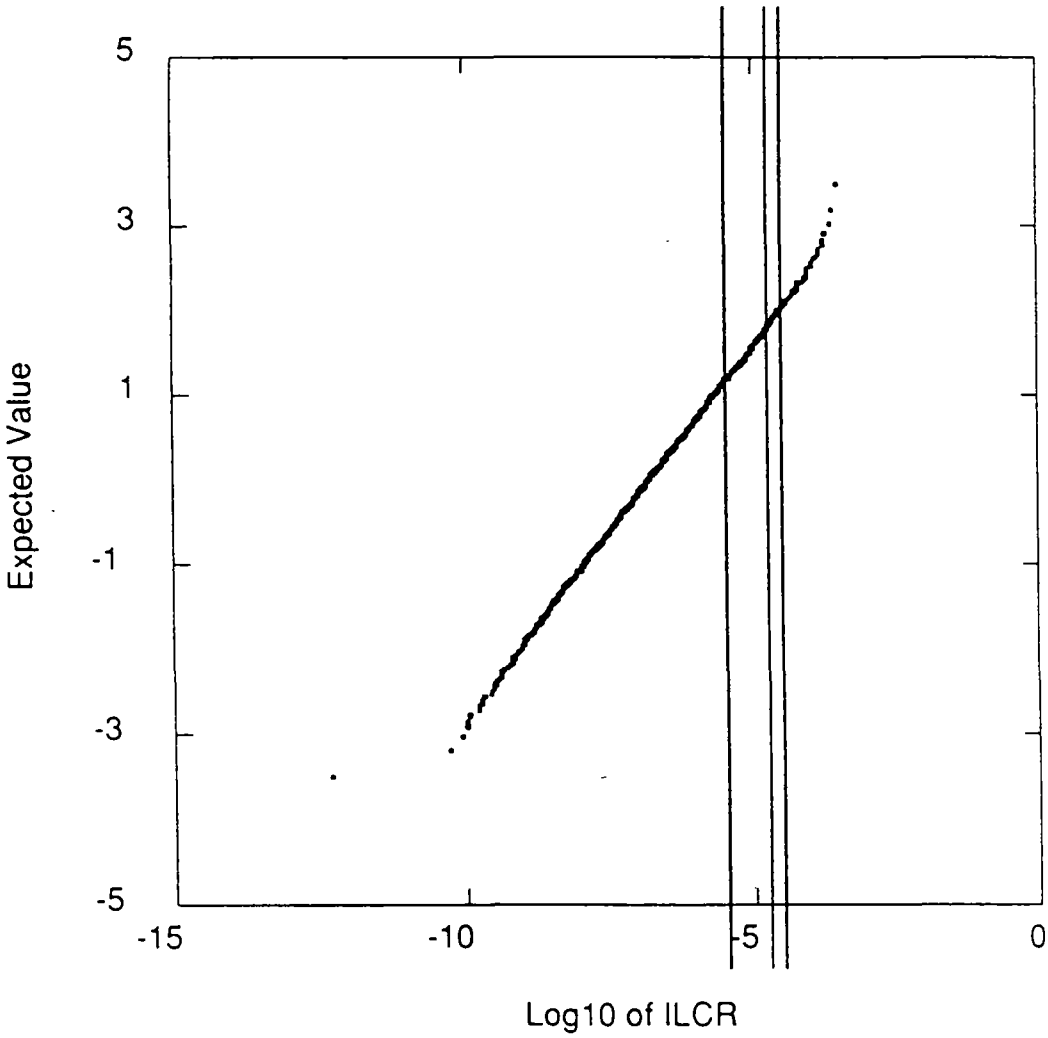


**EXHIBIT 8. Summary Statistics for Different Contributing Factors in Logarithm Space**

| | All Toggles On ..... | Exposure Toggles On ..... | Soil Toggles On ..... | Skin Toggles On ..... | Conc Toggles On ..... | Toxicity Toggles On ..... |
|---|---|---|---|---|---|---|
| Trials | 5,000 | 5,000 | 5,000 | 5,000 | 5,000 | 5,000 |
| Mean | -6.80 | -5 15 | -4 53 | -4 54 | -4 78 | -5 94 |
| Median | -6.81 | -5.14 | -4.53 | -4 53 | -4 78 | -5 93 |
| Mode | -6.84 | -5.19 | -4 52 | -4 52 | -4 76 | -5 88 |
| Standard Deviation | 1.13 | 0.36 | 0.02 | 0 08 | 0 30 | 1 04 |
| Variance | 1.28 | 0 13 | 0.00 | 0 01 | 0 09 | 1 07 |
| Skewness | 0.03 | -0.19 | -0.14 | -0 75 | -0.01 | -0 01 |
| Kurtosis | 2.85 | 2 95 | 3.03 | 4 30 | 3 02 | 2 96 |
| Range Width | 9.84 | 2 57 | 0.12 | 0 71 | 2.22 | 7 81 |
| Minimum | -12.31 | -6 49 | -4 60 | -5 03 | -5 85 | -9 55 |
| Maximum | -2.47 | -3.92 | -4.48 | -4 31 | -3 63 | -1 75 |
| MSE | 0.02 | 0.01 | 0.00 | 0 00 | 0 00 | 0 01 |
| | | | | | | |
| Percentile | | | | | | |
| 0 | -12.31 | -6.49 | -4.60 | -5 03 | -5 85 | -9 55 |
| 5 | -8.70 | -5.74 | -4 56 | -4 68 | -5 26 | -7 64 |
| 10 | -8.27 | -5.60 | -4 55 | -4 64 | -5 16 | -7 27 |
| 15 | -8.00 | -5.52 | -4 55 | -4 62 | -5 08 | -7 01 |
| 20 | -7.79 | -5.45 | -4.54 | -4 60 | -5 03 | -6 81 |
| 25 | -7.58 | -5 39 | -4 54 | -4 58 | -4 98 | -6 64 |
| 30 | -7 40 | -5 33 | -4 54 | -4 57 | -4 93 | -6 48 |
| 35 | -7.25 | -5.28 | -4 54 | -4 56 | -4 89 | -6 34 |
| 40 | -7 09 | -5 23 | -4 53 | -4 55 | -4 85 | -6 20 |
| 45 | -6.95 | -5 18 | -4 53 | -4 54 | -4 82 | -6 07 |
| 50 | -6.81 | -5 14 | -4 53 | -4 53 | -4 78 | -5 93 |
| 55 | -6.67 | -5.09 | -4.53 | -4 52 | -4 74 | -5 80 |
| 60 | -6.53 | -5 04 | -4 52 | -4 51 | -4 70 | -5 67 |
| 65 | -6.37 | -4 99 | -4 52 | -4 50 | -4 66 | -5 54 |
| 70 | -6.21 | -4 94 | -4 52 | -4 49 | -4 62 | -5 39 |
| 75 | -6.04 | -4.89 | -4 52 | -4 48 | -4 58 | -5 23 |
| 80 | -5 83 | -4 84 | -4.52 | -4 47 | -4 53 | -5 06 |
| 85 | -5 61 | -4 78 | -4 51 | -4 46 | -4 47 | -4 86 |
| 90 | -5.35 | -4.69 | -4 51 | -4 44 | -4 40 | -4 61 |
| 95 | -4 96 | -4 58 | -4 50 | -4 42 | -4 29 | -4 23 |
| 100 | -2 47 | -3 92 | -4 48 | -4 31 | -3 63 | -1 75 |

**EXHIBIT 9.**
**Contributions to Overall Uncertainty**

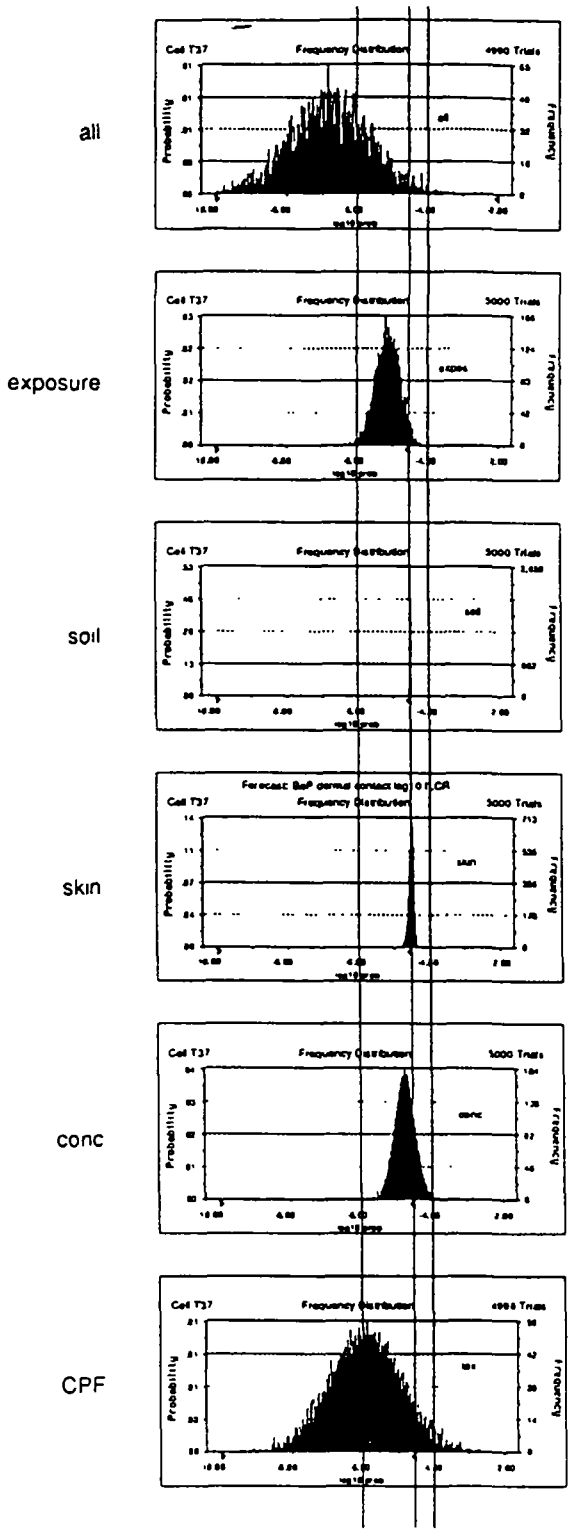



**EXHIBIT 10.**
**Contributions to Overall Uncertainty**

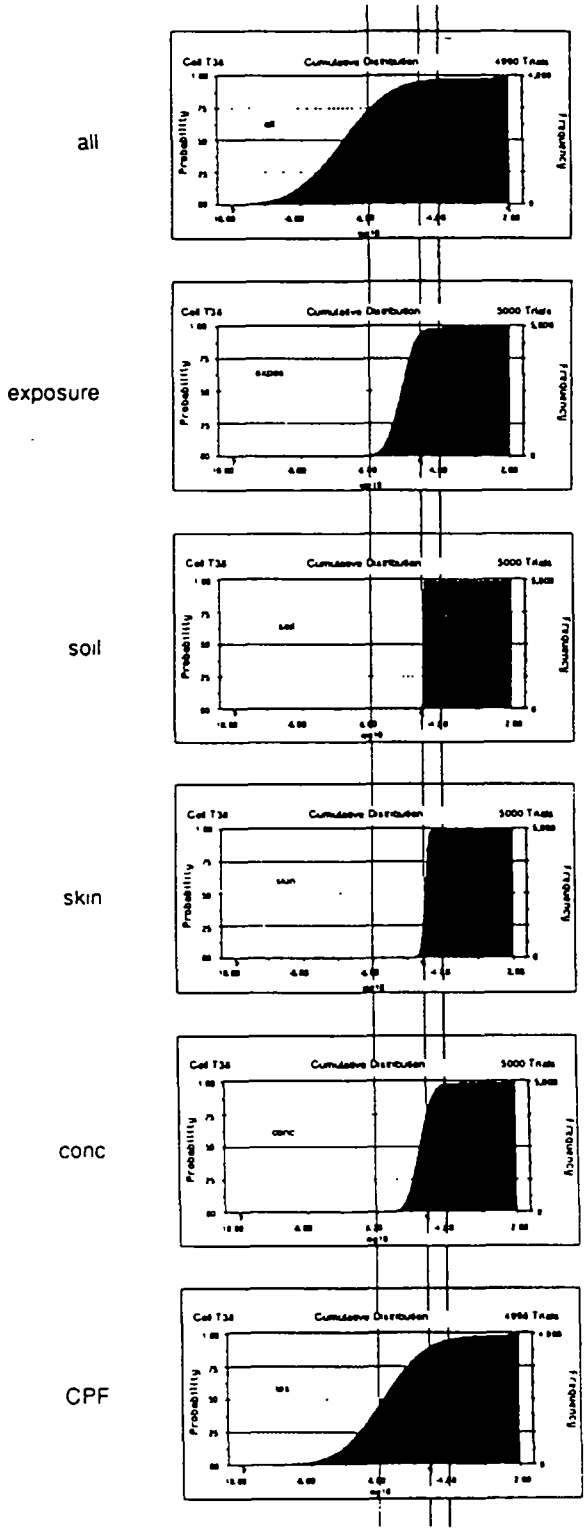


**EXHIBIT 11. Cumulative Distributions for ILCR**

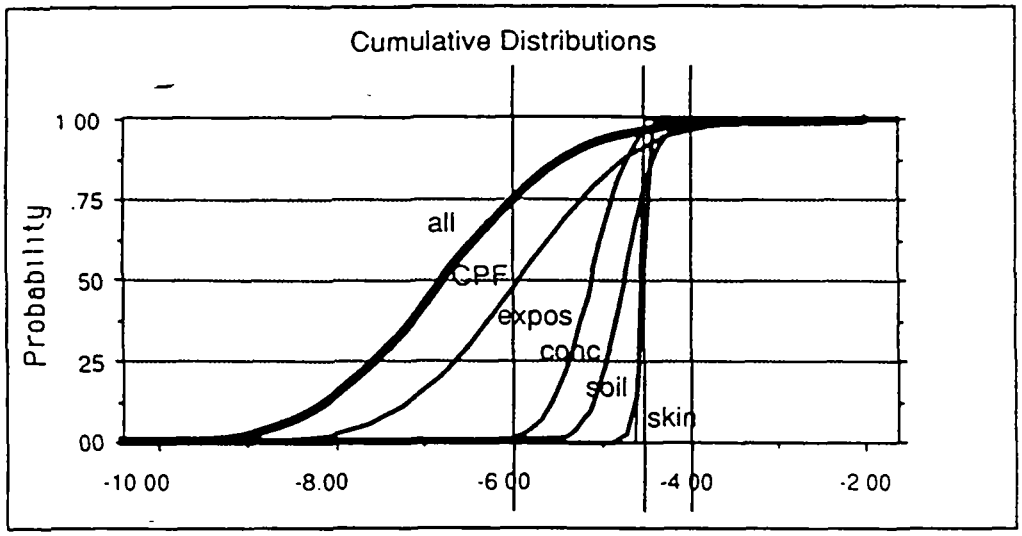



**EXHIBIT 12. Box and Whiskers Diagram Showing Components**

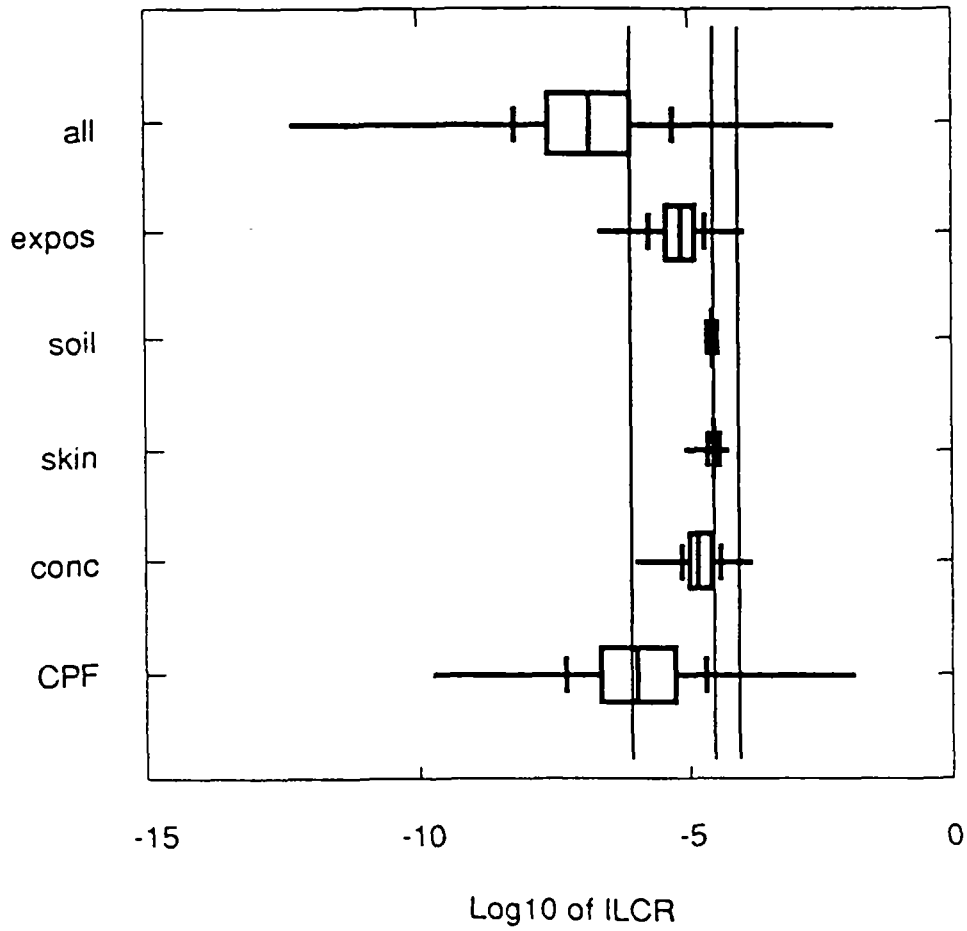
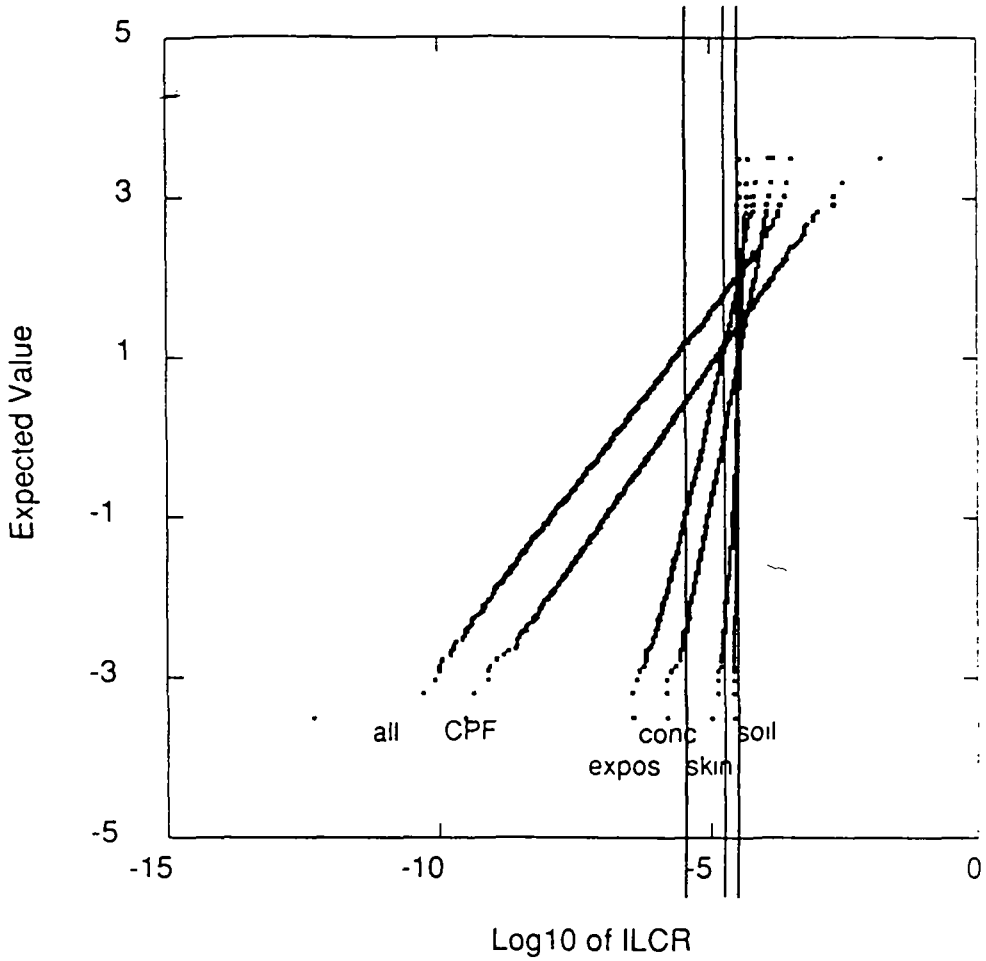


**EXHIBIT 13. Probability Plot of ILCR Showing Components**




## REFERENCES

BENJAMIN, J.R. and CORNELL, C.A. (1970). Probability, Statistics, and Decisions for Civil Engineers. McGraw-Hill, New York, NY.

BROADBENT, S.R. (1956). "Lognormal Approximation to Products and Quotients." Biometrika 43: 404 - 417.

BURMASTER, D.E., THOMPSON, K.M., MENZIE, C.A., CROUCH, E A.C., and McKONE, T.E , (1990). "Monte Carlo Techniques for Quantitative Uncertainty Analysis in Public Health Risk Assessments." In: Proceedings of the National Conference on Hazardous Waste and Hazardous Materials, Washington, DC. HMCRI, Silver Spring, MD. pp. 215-221.

BURMASTER, D.E. and von STACKELBERG, K.E (1989). "Quantitative Uncertainty Analysis in Exposure and Dose-Response Assessments in Public Health Risk Assessments Using Monte Carlo Techniques." In: Proceedings of the National Conference on Hazardous Waste and Hazardous Materials, Washington, DC. HMCRI, Silver Spring, MD. pp. 82-85.

BURMASTER, D.E. and UDELL, E.C. (1990). "A Software Review of Crystal Ball™." Risk Analysis 10(2): 343-345.

CHAMBERS, J.M., CLEVELAND, W.S., KLEINER, B., and TUKEY, P.A. (1983) Graphical Methods for Data Analysis. Wadsworth International Group, Belmont, CA

CLEVELAND, W S. (1985). The Elements of Graphing Data. Wadsworth Advanced Books, Monterey, CA.

CROUCH, E.A.C. (1990). Incorporating Uncertainty into Carcinogenic Risk Assessments for Chemicals, Draft. Cambridge Environmental, Inc., Cambridge, MA.

CROUCH, E.A.C. (1983). "Uncertainties in Interspecies Extrapolations of Carcinogenicity." Environ. Health Perspec. 50: 321-327.

CROUCH, E. and , R. (1981). "Regulation of Carcinogens." Risk Analysis1(1): 47-57.

EBASCO SERVICES, INC. (1990). Final Public Health Evaluation, Halby Chemical Company Site, Wilmington, DE. EPA Contract Number 68-01-7250.

ENVIRON CORPORATION. (1991). Technical Memorandum, Assessment of Off-Site Public Health Risks Posed by a Hypothetical Catastrophic Release from Drummed Materials in Section B of the Rohm & Haas Bristol Landfill.

FENTON, L.F. (1960). "The Sum of Log-Normal Probability Distributions in Scatter Transmission Systems." IRE Transactions on Communications Systems, CS-8: 57 - 67.

FINKEL, A.M. (1990). Confronting Uncertainty in Risk Management, A Guide for Decision-Makers. Resources for the Future, Washington, DC.

GCA CORPORATION (1984). Development of Statistical Distributions or Ranges of Standard Factors Used in Exposure Assessments, Revised Draft Final Report to EPA's Exposure Assessment Group in Washington, GCA/Technology Division, Chapel Hill, NC.

GRAHAM, J.D. and HENRION, M. (1984). "A Probabilistic Analysis of the Passive-Restraint Question, Risk Analysis 4(1): 25 - 30.

IBREKK, H. and MORGAN, M.G. (1983). "Graphical Communication of Uncertain Quantities to Nontechnical People." Risk Analysis 7(4): 519 - 529.

IMAN, R.L. and Conover, W.J. (1982). "A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables." Communications in Statistical Simulation Computation 11(3): 311 - 334.

IMAN, R.L. and DAVENPORT, J.M. (1982). "Rank Correlation Plots for Use with Correlated Input Variables." Communications in Statistical Simulation Computation 11(3): 335 - 360.

MABEY, W.R. et al. (1982). Aquatic Fate Process Data for Organic Priority Pollutants, Final Report, Office of Water Regulations and Standards, EPA/440/4-81-014.

DECISIONEERING, INC. (1991) User Manual for Crystal Ball™ Version 2.0. Decisioneering, Inc., Suite 200, 1727 Conestoga Street, Boulder, CO.

McKONE, T.E. (1990). "Dermal Uptake of Organic Chemicals from a Soil Matrix " Risk Analysis 10(3): 407-419

NATIONAL ACADEMY OF SCIENCES (NAS). (1983). Risk Assessment in the Federal Government: Managing the Process. National Academy Press, Washington, DC.

RUBENSTEIN, R.Y. (1981). Simulation and the Monte Carlo Method, John Wiley & Sons, New York, NY.

SHLYAKHTER, A. (1991). Lognormal Approximations to Probability Distributions of Risk Estimates, Draft Manuscript. Department of Physics, Harvard University, Cambridge, MA.

SMITH, R. (1991). Personal communication with Chief Toxicologist, Region III, U.S. Environmental Protection Agency, Philadelphia, PA.

SPRINGER, M.D. (1979). The Algebra of Random Variables. John Wiley & Sons, New York, NY.

TUFTE, E.R. (1983). The Visual Display of Quantitative Information. Graphics Press, Cheshire, CT.

TUFTE, E.R. (1990). Envisioning Information. Graphics Press, Cheshire, CT

TUKEY, J.W. (1977). Exploratory Data Analysis. Addison Wesley, Reading, MA.

SYSTAT, INC. (1991). User's Manual for Sygraph, Evanston, IL.

U.S. ENVIRONMENTAL PROTECTION AGENCY (EPA). (1990). National Oil and Hazardous Substances Pollution Contingency Plan, revisions to 40 CFR Part 300, published as 55FR866 et seq, 8.

U.S. ENVIRONMENTAL PROTECTION AGENCY (EPA). (1989a). Final Report Exposure Factors Handbook, Office of Health and Environmental Assessment, EPA/600/8-89/043, Washington, DC.

U.S. ENVIRONMENTAL PROTECTION AGENCY (EPA).(1989b). Risk Assessment Guidance for Superfund, Volume I, Human Health Evaluation Manual, Part A, Interim Final, Office of Emergency and Remedial Response, EPA/540/1-89/002, Washington, DC.

U.S. ENVIRONMENTAL PROTECTION AGENCY (EPA). (1989c). EPA's Database documented in: Integrated Risk Information System, Volume I and Electronic Information System, Office of Health and Environmental Assessment, EPA/600/8-86/032a, Washington, DC.

U.S. ENVIRONMENTAL PROTECTION AGENCY (EPA). (1986). Superfund Public Health Evaluation Manual, Office of Emergency and Remedial Response, EPA/540/1-86/060, Washington, DC.

U.S. ENVIRONMENTAL PROTECTION AGENCY (EPA). (1984). Health Effects Assessment of Benzo(a)Pyrene, EPA/540/1-86/022 or PB86-134335, Cincinnati, OH.

received: 5/28/91
accepted: 9/30/91

Risk Analysis, Vol. 12, No 1, 1992

# Monte Carlo Techniques for Quantitative Uncertainty Analysis in Public Health Risk Assessments

Kimberly M. Thompson,[1,2] David E. Burmaster,[1] and Edmund A.C. Crouch[3]

Received October 16, 1990; revised August 1, 1991

Most public health risk assessments assume and combine a series of average, conservative, and worst-case values to derive a conservative point estimate of risk. This procedure has major limitations. This paper demonstrates a new methodology for extended uncertainty analyses in public health risk assessments using Monte Carlo techniques. The extended method begins as do some conventional methods—with the preparation of a spreadsheet to estimate exposure and risk. This method, however, continues by modeling key inputs as random variables described by probability density functions (PDFs). Overall, the technique provides a quantitative way to estimate the probability distributions for exposure and health risks within the validity of the model used. As an example, this paper presents a simplified case study for children playing in soils contaminated with benzene and benzo(a)pyrene (BaP).

KEY WORDS: Risk assessment; Monte Carlo simulation; uncertainty analysis.

## 1. INTRODUCTION

Following guidance published by the U.S. Environmental Protection Agency (EPA), most public health risk assessments assume and combine a series of average, conservative, and worst-case values to derive a point estimate of risk that is presumed to be conservative and protective of public health.[1,2] The *Interim Final Human Health Evaluation Manual,*[3] the most recent guidance document from the EPA headquarters, states:

> . . . Each intake variable in the equation has a range of values. *For Superfund exposure assessments, intake variable values for a given pathway should be selected so that the combination of all intake variables results in an estimate of the reasonable maximum exposure for that pathway.* As defined previously, the reasonable maximum exposure (RME) is the maximum exposure that is reasonably expected to occur at a site. Under this approach, some intake variables may not be at their individual maximum values but when in combination with other variables will result in estimates of RME. . . . (p. 6–19, emphasis in the original)

Unfortunately, the Agency offers no further definition—either qualitative or quantitative—for the key concept of reasonable maximum exposure. The guidance does not address the amount of conservatism which should be used in risk assessment.

The current risk assessment procedures have three major limitations. First, by selecting a combination of moderate, conservative, and worst-case assumptions, risk assessors and risk managers have no way of knowing the degree of conservatism in an assessment. Since current risk assessments generally lack sufficient uncertainty analysis, risk managers and the public may have a difficulat time putting the point estimates into some kind of perspective. Second, by setting the bias high enough to swamp the uncertainty for each of many variables—but not necessarily all the variables—risk assessments may consider scenarios that will rarely (if ever) happen. Third, it is fundamentally meaningless to run

[1] Alceon Corporation, P.O. Box 2669, Cambridge, Massachusetts 02238.
[2] Present address: Harvard School of Public Health, 665 Huntington Avenue, Bldg. I Room G13B, Boston, Massachusetts 02115.
[3] Cambridge Environmental, 58 Charles Street, Cambridge, Massachusetts 02141.



0272-4332/92/0300-0053$06 50/1 

the site, Acme cleared the site and removed the visually stained surface soils. However, in further talks with the city last year, Acme agreed in principle to sell the property for inclusion in the park. Depending on the outcome of a site risk assessment for the surface soils on the site, Acme retains the right to limit the use of the site to activities with little or no soil contact (e.g., a parking lot with concession stands, or a swimming pool with large concrete pavilions).

Since our purpose is to illustrate the use of Monte Carlo simulation, we consider only one of the many scenarios which could be considered for this site. The scenario considers children who would play in the park extension contemplated for the old Acme/Baker property. We assume that the children (from ages 8–18 years) will spend 3 hr per day playing at the park on the site and that they visit the park 1 day per week, 20 weeks per year for 10 years. We make the conservative and simplifying assumption that the children contact the soil enough with their hands and lower arms to have a rate of soil deposition on their skin of ~1 mg/cm$^2$ per day, and to ingest ~50 mg of soil from the site per day. Given the uncertainties inherent in an exposure assessment, this scenario is constructed in accordance with current EPA guidelines and using conservative (or health-protective) assumptions, in the spirit of analyzing the RME case, not the absolutely worst case.

## 3. EXPOSURE MODELS

To estimate health effects for compounds with carcinogenic potential, we first estimate the average daily dose that a person receives in units of milligram of bioavailable chemical per kilogram of body weight per day (mg/(kg·d)), averaged over a 70-year life [abbreviated as the ADD(life)]. The scenario requires two exposure models: (i) incidental ingestion of soil and (ii) dermal contact with soil.

Table I shows the 27 variables and constants in the two exposure models and the two Cancer Potency Factors (CPFs). The first two columns of the table show the name, symbol, and units of the variable or constant. The third column indicates whether the parameter applies to the dermal contact model, the soil ingestion model, or both. The fourth column gives the point estimates for the inputs, and the fifth column shows the parameterized distribution we used for those inputs we chose to vary. The sixth column specifies the sources of each of the point estimates and distributions, and the seventh column gives the location of the point estimate in the distribution. All of the point values are reasonable in the sense that the EPA has or could readily endorse the values for a particular site. Table II shows the exposure models [used to estimate the ADD(life) values] and the risk equations.

### 3.1. Ingestion of Soil

In this simplified case, we consider exposures from the incidental and inadvertent ingestion of contaminated soil (i.e., we include only children who do not exhibit pica). Equation (1) in Table II shows the exposure model used to estimate the ADD(life) for inadvertent ingestion of contaminated soil.

### 3.2. Dermal Contact with Soil

Risk assessments often evaluate exposures from dermal contact with contaminated soils. In 1990, McKone published a new model which estimates the uptake of chemicals from a soil matrix deposited onto the skin surface.[9] In this model, the stratum corneum is the main barrier to uptake, and the amount of chemical which passes through the stratum corneum represents the bioavailable dose. The model depends on scenario specific inputs, soil properties, skin properties, and chemical properties of the soil contaminants. Although both continuous and one-time deposition versions of the model are available, we use the one-time or unit-deposition model in this simplified analysis.

The unit-deposition model derives a Personal Exposure Factor (PEF) which, when multiplied by the concentration of the chemical in the soil, estimates the average daily dose on a day of exposure. Equation (2) in Table II shows the exposure model used to estimate the ADD(life) for dermal contact with contaminated soil. This PEF is averaged over a day of exposure and is a function of 17 variables as shown in Eqs. (3)–(5) in Table II. (Note that Eqs. (3)–(5) are only given to show how the different variables are used in the model. For details about the model, see Ref. 9.)

Since this model requires 17 inputs (and creating or finding 17 different parameterized distributions is an arduous task), we performed a standard sensitivity analysis to identify the most sensitive inputs. By varying each input variable ± 10% from its nominal value while holding all the other inputs constant, we found those variables which have the greatest effect on the output when changed. If distributions for all 17 of the input variables had been available, then we would have performed a

Table II. Exposure Model and Risk Equations[a]

Soil ingestion model used to find the ADD(life)[b]:

ADD (life) =

$$\frac{Cs \quad SIngR \cdot RBA \cdot DpW \cdot WpY \cdot YpL \quad 10^{-6}\ kg/mg}{BW \quad DinY \cdot YinL} \tag{1}$$

Dermal contact with soil model used to find the ADD(life)[b]:

$$ADD(life) = \frac{Cs \cdot PEF \cdot DpW \quad WpY \cdot YpL}{DinY \cdot YinL} \tag{2}$$

where:

$$PEF = \frac{SL \cdot BF \cdot SA \cdot 0.01}{BW}\left(\frac{K_u}{(K_u + K_v)}\right)\left(1 - \exp\left(-\frac{3600\,(\rho_b + 1000 \cdot \Theta + \phi - \Theta)\,(K_u + K_v)\,T}{SL \cdot 0.01}\right)\right) \tag{3}$$

$$K_v = \frac{0.000005 \cdot K_h}{\delta_a\,(4.8 \times 10^{-4}\,\rho_b\,f_{oc}\,K_{ow} + \Theta + K_h\,(\phi - \Theta))} \tag{4}$$

$$\frac{1}{K_u} = \frac{\delta_{skin}\,f_{fat}\,K_{ow}}{D_{water}\,\gamma^{(4/3)}} + \frac{SL \cdot 0.01 \cdot \phi^2\,(4.8 \times 10^{-4}\,\rho_b\,f_{oc}\,K_{ow} + \Theta + K_h\,(\phi - \Theta))}{(\rho_b + 1000 \cdot \Theta + \phi - \Theta)\,((\phi - \Theta)^{3.33}\,D_{air}\,K_h + \Theta^{3.33}\,D_{water})} \tag{5}$$

Equation used to find the ILCR[c]:

$$ILCR = ADD(life) \cdot CPF \tag{6}$$

[a]See Table I for key to symbols.

[b]Average daily dose of a compound, averaged over life during which exposure occurs, in units of mg/(kg·d).

[c]Incremental lifetime cancer risk, the additional probability that a person will develop cancer during lifetime in which exposure occurs (dimensionless probability).

## 4. POINT ESTIMATES AND PARAMETERIZED DISTRIBUTIONS

In this paper, we use three well-known distributions to describe the key model inputs: the normal or Gaussian distribution, the lognormal distribution, and the uniform distribution. We denote random variable $X$ with a normal distribution as $X \sim$ Normal ($\mu$, $\sigma$), where $\mu$ and $\sigma$ represent the arithmetic mean and standard deviation, respectively. Similarly, the lognormal distribution is denoted as $X \sim$ Lognormal ($m$, $s$), where $m$ and $s$ represent the arithmetic mean and standard deviation of the underlying normal distribution, respectively. (The underlying normal distribution is generated by taking the logarithms of the values in the distribution.) Finally, we use the notation $X \sim$ Uniform ($x_1$, $x_2$) to show that the random variable $X$ is distributed uniformly between fixed minimum ($x_1$) and maximum ($x_2$) values.

### 4.1. Chemical Concentrations in the Soils

For this hypothetical site, we synthesize a data set consistent with the site history. We estimate the exposure point concentration for each chemical in the soils as the 95th percentile of the arithmetic mean of the soil data (i.e., 3.39 mg/kg for benzene and 29.49 mg/kg for BaP). Next, following the Monte Carlo framework, we fit lognormal distributions to the synthetic data for each chemical to estimate PDFs for the exposure point concentrations (where $Cs$ represents the concentration of the chemical in the soils on the site in mg/kg): $Cs_{benzene} \sim$ Lognormal (0.84, 0.77) and $Cs_{BaP} \sim$ Lognormal (2.81, 0.68).

### 4.2. Cancer Potency Factors

Because of the assumptions made and the methodology used in their derivation, CPF values estimated from human or animal data are inherently uncertain values. Incorporating uncertainties into risk assessments requires careful consideration of where such uncertainties arise, methods of characterizing those uncertainties, and the results of such methodologies (e.g., the sizes of the uncertainties) in particular cases. There are many potential sources of uncertainty, including the experimental results, the epidemiological model and doses, the interspecies extrapolation, and the route extrapolation. Extending the ideas in earlier publications,[15–17] one author (EC) evaluated the EPA CPFs for benzene and BaP, and estimated the degree to which the EPA values are overly conservative (biased) and uncertain. Based on this information, we parameterize the CPFs for benzene and BaP, for use in quantitative uncertainty analyses, as lognormal distributions conditional on certain modeling assumptions. We assume that extrapolation between animals and humans is unbiased if performed on the basis of body weight. We divide the EPA point estimate by the amount of bias (the factor by which the EPA value overestimates the median) to obtain the median of the distribution. To be consistent with our notation, we find the natural logarithm of this value to describe the distribution. Similarly, we use the natural logarithm of the uncertainty associated with the EPA "standard" value as the standard deviation. The CPFs for benzene and BaP have these distributions (each in units of (mg/(kg·d))$^{-1}$): $CPF_{benzene} \sim$ Lognormal ($-4.33$, 0.67) and $CPF_{BaP} \sim$ Lognormal ($-0.79$, 2.39).

We choose the published EPA ingestion CPFs as the point estimates of the CPFs for benzene and BaP, 2.9E-02 and 11.5 (mg/(kg·d))$^{-1}$, respectively.[18–20] These

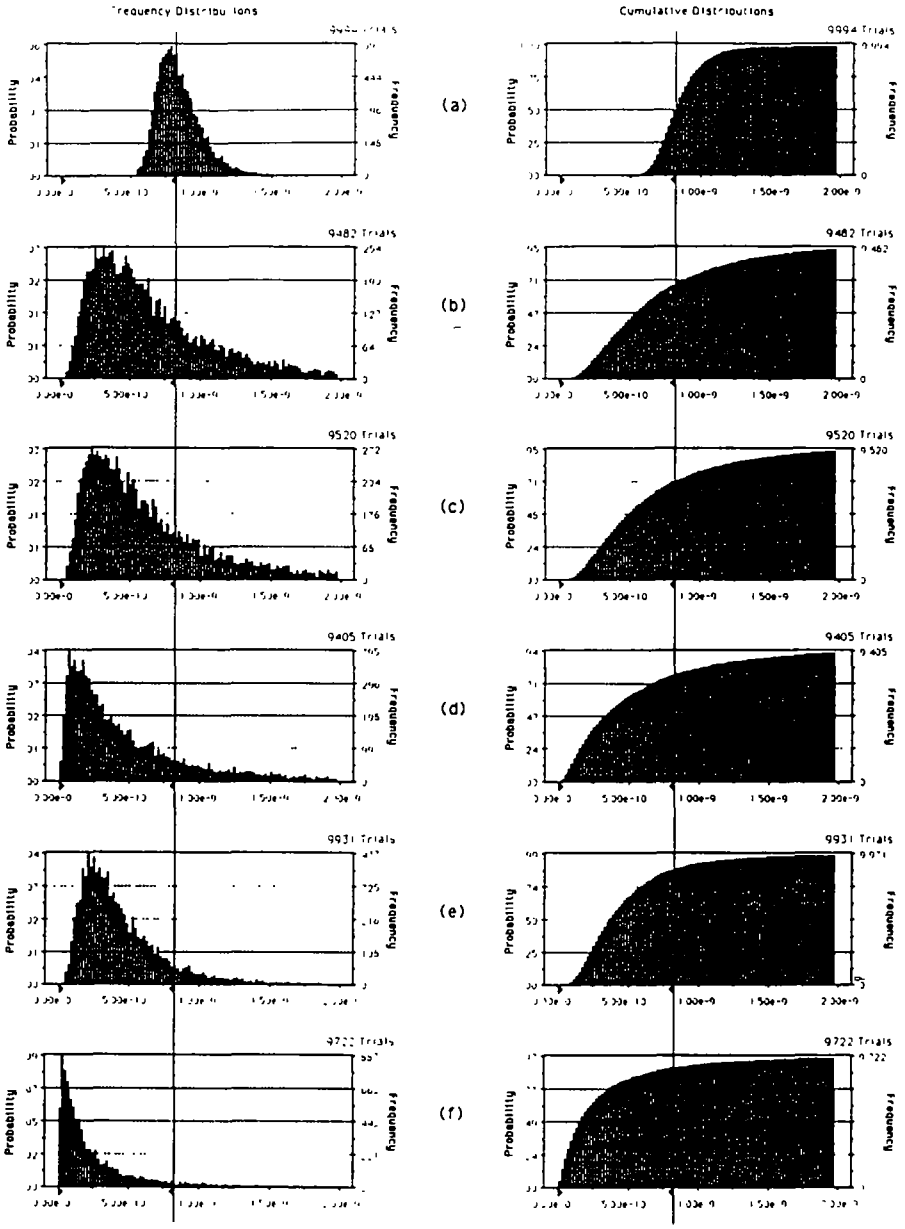



Fig. 1. Frequency and cumulative distributions for ILCR from ingestion of soils contaminated with benzene.

expected, these output distributions have long right tails, high variance, and average values much lower than the point estimates. The locations of the point estimates using all of the exposure variables shift to the 78th percentile for the ingestion of benzene in soil case and they shift to the 94th percentile for the dermal contact with BaP case.

The CPFs for benzene and BaP with the distributions given earlier are group IV random variables. Figures 1e and 2e show the distributions for the five measures of risk. As expected, we see dramatic shifts in the distributions toward values lower than the point estimates. For each of the two pathways, the point estimates fall at the 88th percentile for benzene and the 91st percentile for BaP. These simulations demonstrate the amount of conservatism built into the CPFs.

Finally, Figs. 1f and 2f show the distributions using groups I, II, III, and IV random variables in the simu-

Table III. Summary Statistics for Distributions Shown in Figures 1 and 2

| Stochastic variable groups: | I | II | III | I,II,III | IV | I,II,III,IV |
|---|---|---|---|---|---|---|
| Benzene soil ingestion ILCR | (a) | (b) | (c) | (d) | (e) | (f) |
| **Statistics:** | | | | | | |
| Point estimate location | 50% | 69% | 72% | 78% | 88% | 90% |
| Mean | 8.51E−10 | 7.61E−10 | 7.17E−10 | 6.64E−10 | 4.65E−10 | 3.74E−10 |
| Median (exact) | 8.21E−10 | 5.65E−10 | 5.16E−10 | 3.58E−10 | 3.67E−10 | 1.61E−10 |
| Mode | 5.15E−10 | 3.54E−11 | 3.34E−11 | 6.22E−12 | 2.20E−11 | 1.25E−12 |
| **Percentile:** | | | | | | |
| 0% | 5.15E−10 | 3.54E−11 | 3.34E−11 | 6.22E−12 | 2.20E−11 | 1.25E−12 |
| 5% | 6.39E−10 | 1.63E−10 | 1.42E−10 | 5.45E−11 | 1.22E−10 | 1.91E−11 |
| 10% | 6.69E−10 | 2.11E−10 | 1.86E−10 | 8.15E−11 | 1.58E−10 | 3.05E−11 |
| 15% | 6.95E−10 | 2.55E−10 | 2.23E−10 | 1.09E−10 | 1.84E−10 | 4.21E−11 |
| 20% | 7.14E−10 | 2.98E−10 | 2.61E−10 | 1.37E−10 | 2.10E−10 | 5.46E−11 |
| 25% | 7.33E−10 | 3.39E−10 | 3.00E−10 | 1.66E−10 | 2.35E−10 | 6.78E−11 |
| 30% | 7.51E−10 | 3.78E−10 | 3.38E−10 | 1.95E−10 | 2.58E−10 | 8.21E−11 |
| 35% | 7.68E−10 | 4.25E−10 | 3.80E−10 | 2.29E−10 | 2.85E−10 | 9.91E−11 |
| 40% | 7.86E−10 | 4.70E−10 | 4.21E−10 | 2.65E−10 | 3.11E−10 | 1.18E−10 |
| 45% | 8.03E−10 | 5.14E−10 | 4.68E−10 | 3.07E−10 | 3.39E−10 | 1.39E−10 |
| 50% | 8.21E−10 | 5.65E−10 | 5.16E−10 | 3.58E−10 | 3.67E−10 | 1.61E−10 |
| 55% | 8.39E−10 | 6.21E−10 | 5.70E−10 | 4.13E−10 | 4.01E−10 | 1.87E−10 |
| 60% | 8.60E−10 | 6.84E−10 | 6.32E−10 | 4.77E−10 | 4.38E−10 | 2.24E−10 |
| 65% | 8.82E−10 | 7.53E−10 | 6.98E−10 | 5.49E−10 | 4.79E−10 | 2.65E−10 |
| 70% | 9.08E−10 | 8.41E−10 | 7.84E−10 | 6.46E−10 | 5.29E−10 | 3.16E−10 |
| 75% | 9.34E−10 | 9.51E−10 | 8.93E−10 | 7.49E−10 | 5.82E−10 | 3.80E−10 |
| 80% | 9.69E−10 | 1.08E−9 | 1.02E−9 | 8.98E−10 | 6.49E−10 | 4.76E−10 |
| 85% | 1.01E−9 | 1.25E−9 | 1.20E−9 | 1.12E−9 | 7.35E−10 | 6.14E−10 |
| 90% | 1.06E−9 | 1.52E−9 | 1.47E−9 | 1.47E−9 | 8.80E−10 | 8.59E−10 |
| 95% | 1.16E−9 | 2.03E−9 | 1.97E−9 | 2.21E−9 | 1.13E−9 | 1.38E−9 |
| 100% | 3.04E−9 | 1.10E−8 | 8.47E−9 | 1.95E−8 | 5.14E−9 | 2.29E−8 |
| BaP dermal contact ILCR | (a) | (b) | (c) | (d) | (e) | (f) |
| **Statistics:** | | | | | | |
| Point estimate location | 51% | 80% | 94% | 94% | 91% | 97% |
| Mean | 3.06E−5 | 2.12E−5 | 1.50E−5 | 1.09E−5 | 2.10E−5 | 7.72E−6 |
| Median (exact) | 2.94E−5 | 1.68E−5 | 1.30E−5 | 7.22E−6 | 1.18E−6 | 2.87E−7 |
| Mode (exact) | 7.94E−6 | 1.25E−6 | 1.55E−6 | 1.90E−7 | 8.44E−11 | 1.13E−11 |
| **Percentile:** | | | | | | |
| 0% | 7.94E−6 | 1.25E−6 | 1.55E−6 | 1.90E−7 | 8.44E−11 | 1.13E−11 |
| 5% | 1.80E−5 | 5.48E−6 | 5.48E−6 | 1.61E−6 | 2.37E−8 | 4.30E−9 |
| 10% | 2.03E−5 | 6.95E−6 | 5.57E−6 | 2.24E−6 | 5.73E−8 | 1.06E−8 |
| 15% | 2.18E−5 | 8.28E−6 | 7.60E−6 | 2.82E−6 | 1.04E−7 | 2.01E−8 |
| 20% | 2.32E−5 | 9.42E−6 | 8.38E−6 | 3.36E−6 | 1.64E−7 | 3.14E−8 |
| 25% | 2.43E−5 | 1.06E−5 | 9.17E−6 | 3.90E−6 | 2.43E−7 | 4.89E−8 |
| 30% | 2.54E−5 | 1.17E−5 | 9.91E−6 | 4.48E−6 | 3.46E−7 | 7.19E−8 |
| 35% | 2.64E−5 | 1.30E−5 | 1.07E−5 | 5.04E−6 | 4.78E−7 | 1.04E−7 |
| 40% | 2.74E−5 | 1.41E−5 | 1.14E−5 | 5.70E−6 | 6.51E−7 | 1.43E−7 |
| 45% | 2.84E−5 | 1.54E−5 | 1.22E−5 | 6.40E−6 | 8.82E−7 | 2.05E−7 |
| 50% | 2.94E−5 | 1.68E−5 | 1.30E−5 | 7.22E−6 | 1.18E−6 | 2.87E−7 |
| 55% | 3.05E−5 | 1.83E−5 | 1.39E−5 | 8.08E−6 | 1.59E−6 | 3.85E−7 |
| 60% | 3.16E−5 | 2.00E−5 | 1.49E−5 | 9.07E−6 | 2.21E−6 | 5.33E−7 |
| 65% | 3.28E−5 | 2.18E−5 | 1.60E−5 | 1.02E−5 | 3.02E−6 | 7.61E−7 |
| 70% | 3.40E−5 | 2.40E−5 | 1.72E−5 | 1.17E−5 | 4.13E−6 | 1.08E−6 |
| 75% | 3.55E−5 | 2.66E−5 | 1.85E−5 | 1.35E−5 | 6.02E−6 | 1.59E−6 |
| 80% | 3.71E−5 | 2.96E−5 | 2.03E−5 | 1.56E−5 | 8.81E−6 | 2.49E−6 |
| 85% | 3.93E−5 | 3.37E−5 | 2.24E−5 | 1.84E−5 | 1.39E−5 | 3.90E−6 |
| 90% | 4.22E−5 | 3.97E−5 | 2.56E−5 | 2.31E−5 | 2.50E−5 | 7.50E−6 |
| 95% | 4.70E−5 | 5.16E−5 | 3.11E−5 | 3.20E−5 | 5.98E−5 | 1.88E−5 |
| 100% | 1.22E−4 | 2.44E−4 | 1.04E−4 | 2.50E−4 | 1.60E−2 | 1.19E−2 |

Uncertainty and Sensitivity Analysis in Public Health Risk Assessments at Hazardous Waste Sites Using Monte Carlo Techniques in a Spreadsheet," *Proc Natl. Conf. on Hazardous Waste and Hazardous Materials* (Washington, D.C., HMCRI, Silver Spring, November 1988), pp. 550–556

8. A. M. Finkel, *Confronting Uncertainty in Risk Management, A Guide for Decision-Makers* (Resources for the Future, Washington, D C., 1990).
9. T. E. McKone, "Dermal Uptake of Organic Chemicals from a Soil Matrix," *Risk Analysis*, **10**, 407–419 (1990).
10. The authors of this paper, based on other sources (cited). For details about a particular distribution, contact authors.
11. GCA Corporation, *Development of Statistical Distributions or Ranges of Standard Factors Used in Exposure Assessments* (revised draft final report to U.S. EPA's Exposure Assessment Group in Washington, GCA/Technology Division, Chapel Hill, North Carolina, September 1984).
12. P. K. LaGoy, "Estimated Soil Ingestion Rates for Use in Risk Assessment," *Risk Analysis* **7**, 355–359 (1987).
13. K. M. Thompson and D. E. Burmaster, "Parametric Distributions for Soil Ingestion by Children," *Risk Analysis* **11**, 339–342 (1991).
14. Mabey, W. R. *et al., Aquatic Fate Process Data for Organic Priority Pullutants, Final Report* (Office of Water Regulations and Standards, EPA/440/4–81/014, December 1982).
15. E. A. C. Crouch, R. Wilson, and L. Zeise, "The Risks of Drinking Water," *Water Resources Research* **19**, 1359–1375 (1983).
16. E. A. C. Crouch, "Uncertainties in Interspecies Extrapolations of Carcinogenicity," *Environmental Health Perspectives* **50**, 321–327 (1983).
17. E. A. C. Crouch and R. Wilson, "Regulation of Carcinogens," *Risk Analysis* **1**, 47–57 (1981).
18. U.S. Environmental Protection Agency, US EPA's Database documented in: *Integrated Risk Information System, Volume I and Electronic Information System* (Office of Health and Environmental Assessment, EPA/600/8–86/032a, Washington, D.C., March 1987).
19. U.S. Environmental Protection Agency, *Superfund Public Health Evaluation Manual* (Office of Emergency and Remedial Response, EPA/540/1–86/060, Washington, D.C., October 1986).
20. U.S. Environmental Protection Agency, *Health Effects Assessment of Benzo(a)Pyrene* (EPA/540/1–86/022 or PB86–134335, Cincinnati, September 1984).
21. D. E. Burmaster and E. C. Udell, "A Software Review of Crystal Ball®," *Risk Analysis* **10**, 343–345 (1990).
22. J. S. Salmento, E. S. Rubin, and A. M. Finkel, "A Review of @Risk®," *Risk Analysis* **9**, 255–257 (1989).
23. W. T. Barton, "Response from Palisade Corporation," *Risk Analysis* **9**, 259–260 (1989).

Risk Analysis, Vol 11, No 2, 1991

# Parametric Distributions for Soil Ingestion by Children

Kimberly M. Thompson[1,2] and David E. Burmaster[1]

Received May 4, 1990

This note presents parameterized distributions of estimates of the amount of soil ingested by children based on data collected by Binder *et al.* (1986). Following discussions with Dr. Binder, we modified the Binder study data by using the actual stool weights instead of the 15 g value used in the original study. After testing the data for lognormality, we generated parameterized distributions for use in risk assessment uncertainty analyses such as Monte Carlo simulations.

KEY WORDS: Soil ingestion rates; risk assessment; Monte Carlo simulation, parametric distributions, uncertainty analysis.

## 1. INTRODUCTION

For use in risk assessments, several papers present empirical data and point estimates of the amount of soil ingested by children. In 1987, LaGoy[1] presented an authoritative review of the studies to date and estimated soil ingestion based on age. Although most of the other papers present tables of summary statistics and/or histograms that show large variabilities in the results, none of the papers present enough information for full quantitative uncertainty analysis. To date, no one has presented parameterized distributions which would be useful in Monte Carlo simulations. Using Monte Carlo techniques to estimate both point values *and* the full distributions of the public health risks for a situation make the analyses more informative to risk managers and members of the public because they show where the point estimate falls within the distribution as well as showing the full distributions of risk.[2] However, performing Monte Carlo simulations requires parameterized distributions of each of the key input variables.

Binder *et al.* performed a "diaper study" in 1986, one of the first empirical studies on soil ingestion by children.[3] In the study, the children had an average stool weight of 7.5 g/day, which was half of what previous investigators had measured in other studies.[3,4] Consequently, the authors did not use the actual stool weights of the children in estimating soil ingestion, but instead used 15 g/day as the stool weight for every subject, effectively doubling their estimates of soil ingestion.

In September and October of 1986, Calabrese *et al.*[5] performed another "diaper study" using 8 trace elements (Al, Ba, Mn, Si, Ti, V, Y, and Zr) and a mass-balance approach to account for trace elements ingested in foods and medicines. The authors of that study concluded that not considering ingestion of trace elements in food (particularly Ti and Y) elevates estimates of soil ingestion by factors of between 2 and 6 depending on the trace element. Their findings are consistent with previous studies by Binder *et al.*[3] and Clausing *et al.*[6] if the previous studies are corrected for trace elements ingested in food and medicine. They note that adjusting the fecal weights in the Binder *et al.*[3] study "in retrospect based on [their] data was not justified." In a separate publication, Stanek *et al.*[7] reported the trace element content of the foods and medicines in the Calabrese study; however, these data were not reported in such a way that distributions could be generated.

## 2. METHODS

To fit a parametric distribution to data for soil ingestion by children, we contacted Dr. Binder of the

[1] Alceon Corporation, P.O. Box 2669, Cambridge, Massachusetts 02238.
[2] Current address: Harvard School of Public Health, 665 Huntington Ave., Bldg I Rm G13B; Boston, Massachusetts 02115.



0272-4332/91/0600-0339$06 50/1 

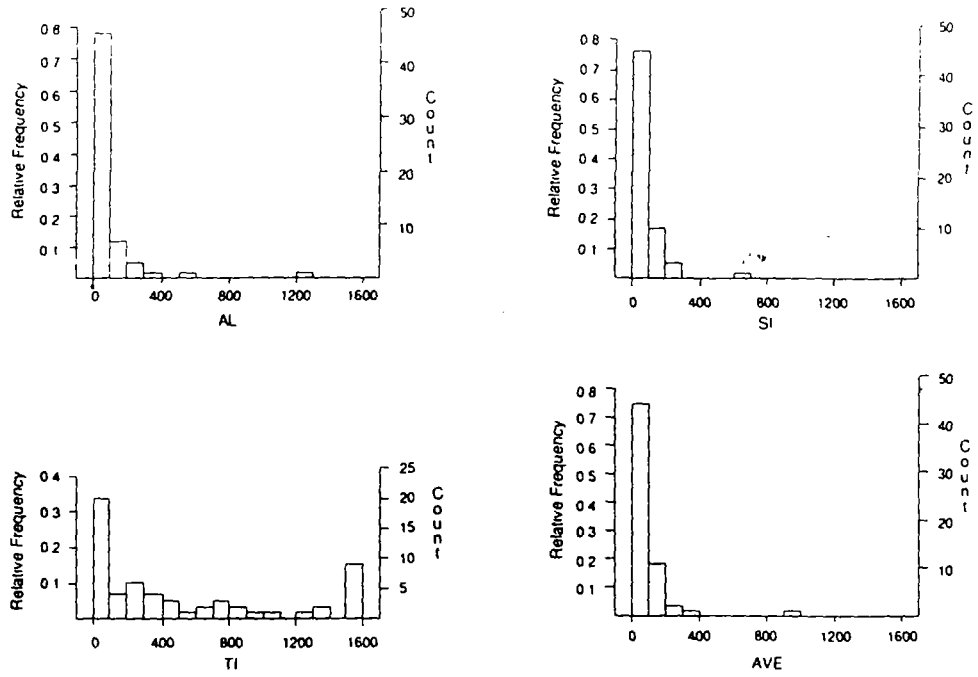



Fig. 1. Histograms of soil ingestion estimates based on Al, Si, Ti, and AVE (in mg/day).

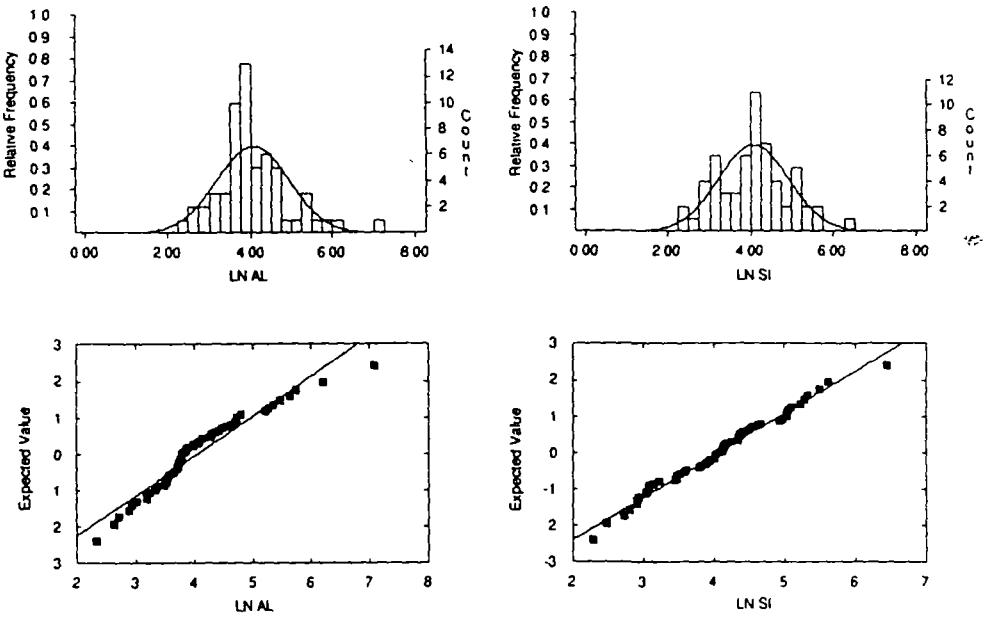



Fig. 2. Histograms and probability plots of soil ingestion estimates based on natural logarithms of Al and Si (in mg/day)

the distribution and its underlying normal (where the underlying normal is found by taking the natural logarithms of the data). For the data, the median and standard deviation are the two parameters generally used to describe the distribution (although we also show the arithmetic mean). For the underlying normal (the natural logarithms of the data), the mean and the standard deviation are shown. As discussed in standard text

published
*Human and Ecological Risk Assessment*
Vol 2 No 2 pp 348-355
1996

# Mathematical Properties of the Risk Equation When Variability is Present

David E. Burmaster
Alceon Corporation
PO Box 382669 Harvard Square Station
Cambridge, MA 02238-2669
tel: 617-864-4300; fax: 617-864-9954
email: deb@Alceon.com

Leslie R. Bloomfield
Stone & Webster Engineering Corporation
245 Summer Street
Boston, MA 02210
tel: 617-589-1765; fax: 617-589-2922
email: leslie.bloomfield@swec.com

Abstract

When random variables are used to represent variability, the risk equation has mathematical properties poorly understood by many risk assessors. Variability represents the heterogeneity in a well-characterized population, usually not reducible through further measurement or study. We follow the lead of most mathematicians in using random variables to represent and analyze variability. To illustrate the issues, we use LogNormal distributions to model variability.

## 1.0 Introduction

When estimating the incremental lifetime cancer risk, R, from an environmental exposure to a single carcinogenic chemical via a single exposure pathway, risk assessors often use equations of this fundamental form:

$$R = \frac{\prod_{i=1}^{I} X_i}{\prod_{j=1}^{J} Y_j} \qquad \text{Eqn 1}$$

where $\prod$ indicates a product over the index. In common practice, risk assessors use point values (i.e., real numbers) for each variable in Eqn 1. Burmaster and Thompson (1995a, b) have discussed the origins and interpretation of Eqn 1 in deterministic risk assessments.

Most risk assessors now agree that all the variables in Eqn 1 contain both (i) variability and/or (ii) uncertainty. In this discussion, variability represents the heterogeneity in a well-characterized population [and is usually not reducible through further measurement or study] while uncertainty represents our

ignorance about a poorly-characterized phenomenon or models [and may be reducible through further measurement or study]. Thus, variability is a property of the natural system under analyst, while uncertainty is a property of the analyst. Here, we focus exclusively on variability -- not because uncertainty is unimportant, but because the introduction of variability alone illustrates the main mathematical points of this discussion.

In the probabilistic paradigm, Eqn 1 remains the fundamental equation of risk assessment (Burmaster & Thompson, 1995a, b). However, in the fully probabilistic framework, each of the variables in Eqn 1 is a positive random variable represented by a probability density function (PDF) or a cumulative distribution function (CDF) (see, e.g., Feller, 1968 & 1971). To emphasize this change in perspective, we re-write Eqn 1 as Eqn 2, with doubly underscored symbols to denote that each variable is now a random variable that expresses variability in a quantity. We also create Eqns 3 and 4, each an alternative and equivalent representation of Eqn 2:

$$\underline{\underline{R}} = \frac{\prod_{i=1}^{I} \underline{\underline{X}}_i}{\prod_{j=1}^{J} \underline{\underline{Y}}_j} \qquad \text{Eqn 2}$$

$$\underline{\underline{R}} = f(\underline{\underline{X}}_i, \underline{\underline{Y}}_j) \quad \text{for } i = 1, \ldots, I \text{ and } j = 1, \ldots, J \qquad \text{Eqn 3}$$

$$\underline{\underline{R}} = \frac{g(\underline{\underline{X}}_i)}{h(\underline{\underline{Y}}_j)} \quad \text{for } i = 1, \ldots, I \text{ and } j = 1, \ldots, J \qquad \text{Eqn 4}$$

In Eqn 4, we use the notation $g(\underline{\underline{X}}_i)$ for the product of random variables in the numerator and the notation $h(\underline{\underline{Y}}_j)$ for the product of random variables in the denominator so we can refer to the numerator and denominator separately as needed. We will continue to denote real variables (point values) without the double underscores. With knowledge of the distributions of all the $\underline{\underline{X}}_i$ and $\underline{\underline{Y}}_j$, an analyst can calculate a closed form expression for the distribution $\underline{\underline{R}}$ in a handful of special cases with independent variables (Springer, 1979). In most practical cases, including those cases with correlated or jointly distributed random variables on the right hand side of the risk equation, the analyst can simulate a numerical approximation to the distribution $\underline{\underline{R}}$ (Rubenstein, 1981; Morgan, 1984).

## 2.0 Background on Two-Parameter LogNormal Distributions

LogNormal distributions with two constant parameters play a central role in expressing variability in human and ecological risk assessment for at least three reasons. First, many physical, chemical, biological, and statistical processes tend to create random variables that follow two-parameter LogNormal distributions for expressing variability (Hattis & Burmaster, 1994). For example, the physical mixing and dilution of one material (say, a miscible or soluble contaminant) into another material (say, surface water in a bay) tends to create non equilibrium concentrations which are LogNormal in character (Ott, 1990; Ott, 1995). Second, when the conditions of the Central Limit Theorem hold, the mathematical process of multiplying a series of random variables will produce a new random variable (the product) which, in the limit, is LogNormal in character, regardless of the distributions from which the input variables arise (Benjamin & Cornell, 1970). Finally, two-parameter LogNormal distributions are self-replicating under multiplication and division, i.e., products and quotients of such LogNormal random variables are themselves distributed lognormally (Aitchison & Brown, 1957; Crow & Shimizu, 1988). All these points apply to Eqns 2, 3, and 4.

The two-parameter LogNormal distribution expressing variability takes its name from the fundamental property that the logarithm of the random variable is distributed according to a Normal or Gaussian distribution (Evans et al, 1993):

$$\ln[\underline{\underline{X}}] \sim N(\mu, \sigma) \qquad \text{Eqn 5}$$

where ln[•] denotes the natural or Napierian logarithm function (base e) and N(•, •) denotes a Normal or Gaussian distribution with two constant parameters, the mean $\mu$ and the standard deviation $\sigma$ (with $\sigma > 0$). In Eqn 5, $\underline{\underline{X}}$ is a two-parameter LogNormal random variable, and $\ln[\underline{\underline{X}}]$ is a Normal random variable. In Eqn 5, $\mu$ is the mean and $\sigma$ is the standard deviation of the distribution for the Normal random variable $\ln[\underline{\underline{X}}]$, not the LogNormal random variable $\underline{\underline{X}}$. Many people say that Eqn 1 represents the LogNormal random variable $\underline{\underline{X}}$ "in logarithmic space." As can be seen in Eqn 5, the random variable $\ln[\underline{\underline{X}}]$ is distributed normally, but the random variable $\underline{\underline{X}}$ is distributed lognormally.

The information coded in Eqn 5 is identical to the information coded in Eqn 6:

$$\underline{\underline{X}} \sim \exp[\, N(\mu, \sigma)\,] \qquad \text{Eqn 6}$$

where exp[•] denotes the exponential function and N(•, •) again denotes the same Normal or Gaussian distribution with the same two constant parameters, mean $\mu$ and standard deviation $\sigma$ (with $\sigma > 0$) as above. In Eqn 6, $\underline{\underline{X}}$ is a two-parameter LogNormal random variable. As earlier, $\mu$ is the mean and $\sigma$ is the standard deviation of the Normal random variable $\ln[\underline{\underline{X}}]$, not the LogNormal random variable $\underline{\underline{X}}$. Many people say that Eqn 6 represents the LogNormal random variable $\underline{\underline{X}}$ "in linear space." When working with Eqn 6 as the representation for a LogNormal random variable $\underline{\underline{X}}$, many people refer to $N(\mu, \sigma)$ as the "underlying Normal distribution" or "the Normal distribution in logarithmic space" as a way to remember its origins.

## 3.0 The Fundamental Risk Equation With All LogNormal Random Variables

### 3.1 The General Case

If all the inputs to the fundamental risk equation, Eqn 2, are independent LogNormal random variables of the form:

$$\underline{\underline{X_i}} \sim \exp[\, N(\mu_i, \sigma_i)\,] \quad \text{for } i = 1, \ldots, I \qquad \text{Eqn 7}$$

$$\underline{\underline{Y_j}} \sim \exp[\, N(\mu_j, \sigma_j)\,] \quad \text{for } j = 1, \ldots, J \qquad \text{Eqn 8}$$

then the distribution of risk is also a LogNormal random variable of the form:

$$\underline{\underline{R}} \sim \exp[\, N(\mu_R, \sigma_R)\,] \qquad \text{Eqn 9}$$

with

$$\mu_R = \sum \mu_i - \sum \mu_j \qquad \text{Eqn 10}$$

$$\sigma_R = \text{Sqrt}\,[\, \sum \sigma^2_i + \sum \sigma^2_j\,] \qquad \text{Eqn 11}$$

with the sums over all the indicated indices. As discussed earlier, LogNormal distributions are self-replicating under multiplication and division.

### 3.2 Working with "High-End" and "Low-End" Values

In 1992, the US Environmental Protection Agency (US EPA) defined the concept of a "high-end" point value for a variable in the numerator of Eqn 2 as a deterministic input to an exposure assessment that falls above the 90$^{th}$ percentile but below the 99.9$^{th}$ percentile of the distribution for the particular random variable (US EPA, 1992). For a variable in the denominator of Eqn 2, one may define a corresponding "low-end" value as falling below the 10$^{th}$ percentile but not below the 0.1$^{th}$ percentile for the particular random variable.

For simplicity of exposition, let us take the 95$^{th}$ percentile as representing a high-end value and the 5$^{th}$ percentile as representing a low-end value of a distribution. Let the notations $\{\underline{\underline{T}}\}_{0.95}$ and $\{\underline{\underline{T}}\}_{0.05}$ and indicate the 95$^{th}$ and 5$^{th}$ percentiles, respectively, of an arbitrary random variable $\underline{\underline{T}}$.

With this notation, when the standard deviations are roughly similar, the high-end value of the numerator of Eqn 4 is considerably smaller than the function of the high-end inputs:

$$\{g(\underline{\underline{X}}_j)\}_{0.95} < g(\{\underline{\underline{X}}_j\}_{0.95}) \quad \text{for } i = 1, \ldots, I \qquad \text{Eqn 12}$$

Similarly, when the standard deviations are roughly similar, the low-end value of the denominator of Eqn 4 is considerably larger than the function of the low-end inputs:

$$\{h(\underline{\underline{Y}}_j)\}_{0.05} > h(\{\underline{\underline{Y}}_j\}_{0.05}) \quad \text{for } j = 1, \ldots, J \qquad \text{Eqn 13}$$

Overall, this means that the high-end value for risk is much, much smaller than the function of the high-end inputs in the numerator and the low-end inputs in the denominator when the standard deviations are roughly similar:

$$\{\underline{\underline{R}}\}_{0.95} \quad << \quad f(\{\underline{\underline{X_i}}\}_{0.95}, \{\underline{\underline{Y_j}}\}_{0.05}) \qquad \text{Eqn 14}$$

$$\text{for } i = 1, \ldots, I \text{ and } j = 1, \ldots, J$$

Most risk assessors now understand this well-documented property of the fundamental risk equation, Eqn 2 (Burmaster & Harris, 1993; Bogen, 1994; Cullen, 1994). This property of the fundamental risk equation does not depend on the use of LogNormal distributions as inputs.

### 3.2 Working with Arithmetic Means

Let the notation $<\underline{\underline{T}}>$ indicate the arithmetic mean (or expected value) of an arbitrary random variable $\underline{\underline{T}}$. For a LogNormal distribution, the arithmetic mean is always greater than the median of the distribution by the factor $\exp[\, 0.5 \cdot \sigma^2_T \,]$. In many practical cases, the arithmetic mean of a LogNormal random variable falls between the 65$^{th}$ and the 80$^{th}$ percentiles of the distribution. However, in certain situations, the arithmetic mean of a LogNormal distribution can exceed the 95$^{th}$ percentile of that distribution.

Some mathematical properties hold in this situation. For independent LogNormal distributions, the arithmetic average of the numerator in Eqn 4 equals the function of the arithmetic averages of the input variables:

$$<g(\underline{\underline{X_i}})> \quad = \quad g(<\underline{\underline{X_i}}>) \qquad \text{for } i = 1, \ldots, I \qquad \text{Eqn 15}$$

Similarly, for independent LogNormal distributions, the arithmetic average of the denominator in Eqn 4 equals the function of the arithmetic averages of the input variables:

$$<h(\underline{\underline{Y_j}})> \quad = \quad h(<\underline{\underline{Y_j}}>) \qquad \text{for } j = 1, \ldots, J \qquad \text{Eqn 16}$$

The results in Eqns 15 and 16 are easy to prove for independent LogNormal distributions, and the results hold generally for other independent random variables from other families of distributions. Some authors use this property as the definition of independence between two random variables.

However, for independent LogNormal distributions, the arithmetic average of risk does not equal the function of the averages of the inputs:

$$\langle \underline{\underline{R}} \rangle \quad \neq \quad f(\langle \underline{\underline{X}}_i \rangle, \langle \underline{\underline{Y}}_j \rangle) \qquad \text{Eqn 17}$$

$$\text{for } i = 1, \ldots, I \text{ and } j = 1, \ldots, J$$

This result in Eqn 17 surprises many people, even though it is easily proved for independent LogNormal distributions. It is true for other families of distributions as well.

### 3.3 Working with Medians

Let the notation $\{\underline{\underline{T}}\}_{0.50}$ indicate the median or 50$^{th}$ percentile of an arbitrary random variable $\underline{\underline{T}}$.

Some mathematical properties hold in this situation. For LogNormal distributions, the median of the numerator in Eqn 4 equals the function of the medians of the input variables:

$$\{g(\underline{\underline{X}}_i)\}_{0.50} \quad = \quad g(\{\underline{\underline{X}}_i\}_{0.50}) \qquad \text{for } i = 1, \ldots, I \qquad \text{Eqn 18}$$

and, the median of the denominator in Eqn 4 equals the function of the medians of the input variables:

$$\{h(\underline{\underline{Y}}_j)\}_{0.50} \quad = \quad h(\{\underline{\underline{Y}}_j\}_{0.50}) \qquad \text{for } j = 1, \ldots, J \qquad \text{Eqn 19}$$

More generally for independent LogNormal distributions, the median risk equals the function of the median inputs to Eqn 3:

$$\{\underline{\underline{R}}\}_{0.50} \quad = \quad f(\{\underline{\underline{X}}_i\}_{0.50}, \{\underline{\underline{Y}}_j\}_{0.50}) \qquad \text{Eqn 20}$$

$$\text{for } i = 1, \ldots, I \text{ and } j = 1, \ldots, J$$

Thus, for independent LogNormal distributions, the median of the function for risk (in Eqns 2, 3, and 4) is the function of the median inputs. Although this result is

not true for independent random variables from other families of distributions, we have found it an excellent approximation in many numerical simulations of Eqns 2, 3, and 4.

### 3.4 Working with Mixed Cases

If we continue to restrict ourselves to independent LogNormal random variables as the inputs to the fundamental risk equation, any of Eqns 2, 3, or 4, then:

- the median of the $\underline{\underline{R}}$ is equal to the function of the medians of the inputs;
- the arithmetic mean of $\underline{\underline{R}}$ is NOT equal to the function of the arithmetic means of the inputs; and
- the 95$^{th}$ percentile of $\underline{\underline{R}}$ is much smaller than the function of (i) the 95$^{th}$ percentiles of all the inputs in the numerator and (ii) the 5$^{th}$ percentiles of all the inputs in the denominator.

Thus, as is exactly true for independent LogNormal distributions and as is approximately true for other independent random variables with longer tails to the right, medians (not averages) are "neutral" and "self replicating" when used as point value inputs to the fundamental risk equation, Eqn 2.

Without doing a full calculation or a full simulation, no one can know the percentile of $\underline{\underline{R}}$ calculated if the inputs to the fundamental risk equation, Eqn 2, include a combination of median values, average values, and high- and low-end values.

Restricting ourselves to the case with independent LogNormal distributions, we see that:

- the use of one or more median values in either the numerator or the denominator does not shift the estimate of R (further) above or (further) below the correct median of $\underline{\underline{R}}$, i.e., median inputs are "neutral" in trying to understand where the value R falls as a percentile of the distribution $\underline{\underline{R}}$;

- the use of one or more average values in the numerator does shift the estimate of R above the correct median of $\underline{\underline{R}}$, i.e., average inputs in the numerator introduce moderate to large (but unknown) amounts of conservatism in trying to understand where the value R falls as a percentile of the distribution $\underline{\underline{R}}$;

- the use of one or more high-end values in the numerator does shift the estimate of R far above the correct median of $\underline{\underline{R}}$, i.e., high-end inputs in the numerator introduce large (but unknown) amounts of conservatism in trying to understand where the value R falls as a percentile of the distribution $\underline{\underline{R}}$; and

- the use of one or more low-end values in the denominator does shift the estimate of R far above the correct median of $\underline{\underline{R}}$, i.e., low-end inputs in the denominator also introduce large (but unknown) amounts of conservatism in trying to understand where the value R falls as a percentile of the distribution $\underline{\underline{R}}$.

Most risk assessors now understand that the introduction of a few high-end values into the numerator or a few low-end values into the denominator of Eqns 1 or 2 can introduce very large amounts of conservatism into the point estimate R (Harris & Burmaster, 1992; Burmaster & Harris, 1993; Bogen, 1994; Cullen, 1994).

Fewer people understand that the introduction of several average values in the numerator of Eqns 1 or 2 can introduce significant amounts -- or even very large amounts -- of conservatism into point estimate R. As an extreme example, if the arithmetic means of three distributions all exceed the $90^{th}$ percentile of the corresponding distribution, the result is obvious. Less obvious, the use of three average values as point values for the corresponding LogNormal random variables can really be the multiplication of three $75^{th}$ percentiles. If these are the only conservative inputs in an equation, these three inputs may multiply to give, in effect, a high-end point value for risk. If these three average values for inputs in the numerator are combined multiplicatively with three high-end values for other inputs in the numerator, the resulting point estimate of risk may be far, far

more conservative than understood just from the combination of the three high-end values along with medians for the other variables.

## 4.0 Conclusions

From this discussion, we draw three main conclusions.

First, without doing a full calculation or a full simulation, no one can know the percentile of $\underline{\underline{R}}$ calculated if the inputs to the fundamental risk equation, Eqn 2, include a combination of median values, average values, and "high end" values.

Second, for independent LogNormal random variables -- and for other independent random variables from other families of distributions with long tails to the right -- the use of one or more medians in the numerator or denominator of Eqns 2, 3, or 4 for input variables does not introduce any compounding conservatisms; in contrast, the use of one or more average values in the numerator of those same equations always introduces multiplicative conservatisms, usually hidden from view and sometimes numerically large.

Third, the simultaneous use of several average values in the numerator (for distributions with long tails to the right) along with several high-end values in the numerator and several low-end values in the denominator can lead to point estimates of risk that fall above the range US EPA uses to set policy.

## Acknowledgments

General Electric and Alceon supported this research.

## References

Aitchison & Brown, 1957
Aitchison, J. and J.A.C. Brown, 1957, The Lognormal Distribution, Cambridge University Press, Cambridge, UK

Benjamin & Cornell, 1970
Benjamin, J.R. and C.A. Cornell, 1970, Probability, Statistics, and Decisions for Civil Engineers, McGraw Hill, New York, NY

Bogen, 1994
Bogen, K.T., 1994, A Note on Compounded Conservatism, Risk Analysis, Volume 14, Number 4, pp 379 - 381

Burmaster & Thompson, 1995, Need
Burmaster, D.E. and K.M. Thompson, 1995, The Need for New Methods to BackCalculate Soil CleanUp Targets in Interval and Probabilistic Risk Assessments, Human and Ecological Risk Assessments, in press

Burmaster & Thompson, 1995, BackCalculating
Burmaster, D.E. and K.M. Thompson, 1995, BackCalculating CleanUp Targets in Probabilistic Risk Assessments When the Acceptability of Cancer Risk is Defined Under Different Risk Management Policies, Human and Ecological Risk Assessments, in press

Burmaster & Harris, 1993, The Magnitude of Compounding Conservatisms in Superfund Risk Assessments, Risk Analysis, Volume 13, Number 2, pp 131 - 143

Crow & Shimizu, 1988
Crow, E.L. and K. Shimizu, Eds., 1988, Lognormal Distributions, Theory and Applications, Marcel Dekker, New York, NY

Cullen, 1994
Cullen, A.C., 1994, Measures of Compounding Conservatism in Probabilistic Risk Assessment, Risk Analysis, Volume 14, Number 4, pp 389 - 393

Evans et al, 1993
Evans, M., N. Hastings, and B. Peacock, 1993, Statistical Distributions, Second Edition, John Wiley & Sons, New York, NY

Feller, 1968 & 1971
Feller, W., 1968 and 1971, An Introduction to Probability Theory and Its Applications, Volumes I and II, John Wiley, New York, NY

Hattis & Burmaster, 1994
Hattis, D.B. and D.E. Burmaster, 1994, Assessment of Variability and Uncertainty Distributions for Practical Risk Assessments, Risk Analysis, Volume 14, Number 5, pp 713 - 730

Morgan, 1984
Morgan, J.T.M., 1984, Elements of Simulation, Chapman and Hall, London, UK

Ott, 1995
Ott, W.R., 1995, Environmental Statistics and Data Analysis, Lewis Publishers, Boca Raton, FL

Ott, 1990
Ott, W.R., 1990, A Physical Explanation of the Lognormality of Pollutant Concentrations, Journal of the Air and Waste Management Association, Volume 40, pp 1378 et seq.

Rubinstein, 1981
Rubinstein, R.Y., 1981, Simulation and the Monte Carlo Method, John Wiley & Sons, New York, NY

Springer, 1979
Springer, M.D., 1979, The Algebra of Random Variables, John Wiley & Sons, New York, NY

US EPA, 1992, Exposure
US Environmental Protection Agency, Guidelines for Exposure Assessment, 57 Federal Register, pp 22888 et seq., 29 May 1992

# Using Beta Distributions Efficiently in A Probabilistic Exposure Assessment

David E. Burmaster
Alceon Corporation
PO Box 382669, Cambridge, MA 02238-2669
tel: 617-864-4300; fax: 617-864-9954
deb@Alceon.com

## Abstract

In a probabilistic exposure assessment, an analyst must often develop a probability distribution to represent a random variable that has a fixed minimum and a fixed maximum. This manuscript shows how to fit a two-parameter Beta distribution to a sample data set, and then shows how to fit a constrained four-parameter Beta4 distribution to the same data -- thereby improving the fit and speeding the simulation by a factor of 5.

## 1.0 Introduction

In a probabilistic exposure assessment, an analyst must often develop a probability distribution to represent a random variable that has a fixed minimum and a fixed maximum. For example, a child may play in a park from 0 to 7 day/week. Similarly, the fraction of skin in contact with soils may range from 0 to 1, while the fraction of a nutrient or a toxicant absorbed in the gut may range from 0 to 1.

Beta distributions (of the first kind) (Mood et al, 1974; Evans et al, 1993) have several useful properties that an analyst may exploit in a probabilistic exposure assessment. First, a two-parameter Beta distribution can assume a wide variety of shapes, depending on the values taken by the two parameters. Second, a two-parameter Beta variate has a fixed minimum (zero) and a fixed maximum (one). Third, a two-parameter

The random variable $\underline{X}$ has this probability density function (PDF):

$$f_X(x) = \frac{x^{a-1} \bullet (1 - x)^{b-1}}{\beta(a, b)} \qquad \text{Eqn 2}$$

where the denominator is the beta function, $\beta(a, b) = \int_0^1 u^{a-1} (1-u)^{b-1} du$ (Abramowitz & Stegun, 1964). The first two central moments of this distribution are, respectively, its expected value and its variance:

$$E[\underline{X}] = \frac{a}{a + b} \qquad \text{Eqn 3}$$

$$V[\underline{X}] = \frac{a \bullet b}{(a + b)^2 (a + b + 1)} \qquad \text{Eqn 4}$$

While the Beta distribution is extremely flexible over the support [ 0, 1 ] and finds many uses in statistics, two of its special properties are key in exposure assessment. First, when a > 1 and b > 1, the distribution has one mode:

$$\text{Mode}[\underline{X}] = \frac{a - 1}{a + b - 2} \qquad \text{Eqn 5}$$

Second, Beta variates are much easier and faster to simulate than when a and b are integers. [EndNote 2]. Further, the speed of the algorithm commonly used to simulate a Beta variate with integer parameters decreases as the sum (a + b) increases (Evans et al, 1993).

Exploratory Data Analysis: After exploratory data analysis to visualize the patterns in the data (Tukey, 1977; Chambers et al, 1983; Cleveland, 1985), we decided to fit a two-parameter Beta distribution to the data.

and these central moments:

$$E[\underline{Y}] = c \bullet E[\underline{X}] \qquad \text{Eqn 9}$$

$$V[\underline{Y}] = c^2 \bullet V[\underline{X}] \qquad \text{Eqn 10}$$

Again, for $a > 1$ and $b > 1$, the Beta4 distribution has one mode:

$$\text{Mode}[\underline{Y}] = c \bullet \text{Mode}[\underline{X}] + d \qquad \text{Eqn 11}$$

Finally, when $c = 1 - d$ and $0 \le d < 1$, the random variate $\underline{Y}$ has support [d, 1].

Exploratory Data Analysis: After further exploratory data analysis, we decided to fit a Beta4 distribution with a and b constrained to integers for speed in simulation.

Application: First, we fit the Beta4 distribution to the data by maximizing the loglikelihood function (Edwards, 1992) in Mathematica™ subject to the constraints: $a = 4$, $b = 1$, and $c = 1 - d$. In Figure 2, a solid line depicts the CDF for this fitted Beta4 distribution, $\underline{Y}_1 \sim \text{Beta4}[\, y \mid \hat{a} = 4, \hat{b} = 1, \hat{c} = 0.397303, \hat{d} = 0.602697 \,]$, plotted against the empirical CDF for the data. The maximum of the loglikelihood function for $\underline{Y}_1$ is 20.24, indicating that this constrained four-parameter Beta4 distribution fits the data better than the best two-parameter Beta distribution. For comparison, the dashed line in Figure 2 depicts the CDF for $\underline{X}_2$.

## 5.0 Discussion

Figure 3 shows the excellent correspondence between the best-fit two-parameter Beta distribution and the best-fit constrained four-parameter Beta4 distribution as a QQ-plot

## References

Abramowitz & Stegun, 1964
Abramowitz, M. and I.A. Stegun, Eds, 1964, Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables, National Bureau of Standards, Applied Mathematics Series Number 55, Issued June 1964, Tenth Printing with corrections in December 1972, Superintendent of Documents, US Government Printing Office, Washington, DC

Chambers et al, 1983
Chambers, J.M., W.S. Cleveland, B. Kleiner, and P.A. Tukey, 1983, Graphical Methods for Data Analysis, Wadsworth International Group, Belmont, CA, and Duxbury Press, Boston, MA

Cleveland, 1985
Cleveland, W.S., 1985, The Elements of Graphing Data, Wadsworth Advanced Books, Monterey, CA

Edwards, 1992
Edwards, A.W.F., 1992, Likelihood, Expanded Edition, John Hopkins University Press, Baltimore, MD

Evans et al, 1993
Evans, M., N. Hastings, and J.B. Peacock, 1993, Statistical Distributions; Second Edition, John Wiley & Sons, Inc., New York, NY

Magee et al, 1996
Magee, B.H., P.D. Anderson, and D.E. Burmaster, 1996, Absorption Adjustment Factor Distributions (AAF) Distributions for Polycyclic Aromatic Hydrocarbons (PAHs), Human and Ecological Risk Assessment, in review

Mood et al, 1974
Mood, A.M., F.A. Graybill, and D.C. Boes, 1974, Introduction to the Theory of Statistics, Third Edition, McGraw Hill, New York, NY

Tukey, 1977
Tukey, J.W., 1977, Exploratory Data Analysis, Addison-Wesley, Reading, MA

Wickham-Jones, 1994
Wickham-Jones, T., 1994, Mathematica Graphics, Techniques & Applications, Springer-Verlag, Telos, Santa Clara, CA

Wolfram, 1991
Wolfram, S., 1991, Mathematica, A System for Doing Mathematics by Computer, Second Edition, Addison-Wesley, Redwood City, CA

# Using Beta Distributions Efficiently in A Probabilistic Exposure Assessment

David E. Burmaster
Alceon Corporation
PO Box 382669, Cambridge, MA 02238-2669
tel: 617-864-4300; fax: 617-864-9954
deb@Alceon.com

## Abstract

In a probabilistic exposure assessment, an analyst must often develop a probability distribution to represent a random variable that has a fixed minimum and a fixed maximum. This manuscript shows how to fit a two-parameter Beta distribution to a sample data set, and then shows how to fit a constrained four-parameter Beta4 distribution to the same data -- thereby improving the fit and speeding the simulation by a factor of 5.

## 1.0 Introduction

In a probabilistic exposure assessment, an analyst must often develop a probability distribution to represent a random variable that has a fixed minimum and a fixed maximum. For example, a child may play in a park from 0 to 7 day/week. Similarly, the fraction of skin in contact with soils may range from 0 to 1, while the fraction of a nutrient or a toxicant absorbed in the gut may range from 0 to 1.

Beta distributions (of the first kind) (Mood et al, 1974; Evans et al, 1993) have several useful properties that an analyst may exploit in a probabilistic exposure assessment. First, a two-parameter Beta distribution can assume a wide variety of shapes, depending on the values taken by the two parameters. Second, a two-parameter Beta variate has a fixed minimum (zero) and a fixed maximum (one). Third, a two-parameter

The random variable $\underline{X}$ has this probability density function (PDF):

$$f_X(x) = \frac{x^{a-1} \bullet (1 - x)^{b-1}}{\beta(a, b)} \qquad \text{Eqn 2}$$

where the denominator is the beta function, $\beta(a, b) = \int_0^1 u^{a-1} (1-u)^{b-1} du$ (Abramowitz & Stegun, 1964). The first two central moments of this distribution are, respectively, its expected value and its variance:

$$E[\underline{X}] = \frac{a}{a + b} \qquad \text{Eqn 3}$$

$$V[\underline{X}] = \frac{a \bullet b}{(a + b)^2 (a + b + 1)} \qquad \text{Eqn 4}$$

While the Beta distribution is extremely flexible over the support [ 0, 1 ] and finds many uses in statistics, two of its special properties are key in exposure assessment. First, when a > 1 and b > 1, the distribution has one mode:

$$\text{Mode}[\underline{X}] = \frac{a - 1}{a + b - 2} \qquad \text{Eqn 5}$$

Second, Beta variates are much easier and faster to simulate than when a and b are integers. [EndNote 2]. Further, the speed of the algorithm commonly used to simulate a Beta variate with integer parameters decreases as the sum (a + b) increases (Evans et al, 1993).

Exploratory Data Analysis: After exploratory data analysis to visualize the patterns in the data (Tukey, 1977; Chambers et al, 1983; Cleveland, 1985), we decided to fit a two-parameter Beta distribution to the data.

and these central moments:

$$E[\underline{Y}] = c \bullet E[\underline{X}] \qquad \text{Eqn 9}$$

$$V[\underline{Y}] = c^2 \bullet V[\underline{X}] \qquad \text{Eqn 10}$$

Again, for $a > 1$ and $b > 1$, the Beta4 distribution has one mode:

$$\text{Mode}[\underline{Y}] = c \bullet \text{Mode}[\underline{X}] + d \qquad \text{Eqn 11}$$

Finally, when $c = 1 - d$ and $0 \le d < 1$, the random variate $\underline{Y}$ has support [d, 1].

Exploratory Data Analysis: After further exploratory data analysis, we decided to fit a Beta4 distribution with a and b constrained to integers for speed in simulation.

Application: First, we fit the Beta4 distribution to the data by maximizing the loglikelihood function (Edwards, 1992) in Mathematica™ subject to the constraints: $a = 4$, $b = 1$, and $c = 1 - d$. In Figure 2, a solid line depicts the CDF for this fitted Beta4 distribution, $\underline{Y}_1 \sim \text{Beta4}[\, y \mid \hat{a} = 4, \hat{b} = 1, \hat{c} = 0.397303, \hat{d} = 0.602697 \,]$, plotted against the empirical CDF for the data. The maximum of the loglikelihood function for $\underline{Y}_1$ is 20.24, indicating that this constrained four-parameter Beta4 distribution fits the data better than the best two-parameter Beta distribution. For comparison, the dashed line in Figure 2 depicts the CDF for $\underline{X}_2$.

## 5.0 Discussion

Figure 3 shows the excellent correspondence between the best-fit two-parameter Beta distribution and the best-fit constrained four-parameter Beta4 distribution as a QQ-plot

## References

Abramowitz & Stegun, 1964
Abramowitz, M. and I.A. Stegun, Eds, 1964, Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables, National Bureau of Standards, Applied Mathematics Series Number 55, Issued June 1964, Tenth Printing with corrections in December 1972, Superintendent of Documents, US Government Printing Office, Washington, DC

Chambers et al, 1983
Chambers, J.M., W.S. Cleveland, B. Kleiner, and P.A. Tukey, 1983, Graphical Methods for Data Analysis, Wadsworth International Group, Belmont, CA, and Duxbury Press, Boston, MA

Cleveland, 1985
Cleveland, W.S., 1985, The Elements of Graphing Data, Wadsworth Advanced Books, Monterey, CA

Edwards, 1992
Edwards, A.W.F., 1992, Likelihood, Expanded Edition, John Hopkins University Press, Baltimore, MD

Evans et al, 1993
Evans, M., N. Hastings, and J.B. Peacock, 1993, Statistical Distributions; Second Edition, John Wiley & Sons, Inc., New York, NY

Magee et al, 1996
Magee, B.H., P.D. Anderson, and D.E. Burmaster, 1996, Absorption Adjustment Factor Distributions (AAF) Distributions for Polycyclic Aromatic Hydrocarbons (PAHs), Human and Ecological Risk Assessment, in review

Mood et al, 1974
Mood, A.M., F.A. Graybill, and D.C. Boes, 1974, Introduction to the Theory of Statistics, Third Edition, McGraw Hill, New York, NY

Tukey, 1977
Tukey, J.W., 1977, Exploratory Data Analysis, Addison-Wesley, Reading, MA

Wickham-Jones, 1994
Wickham-Jones, T., 1994, Mathematica Graphics, Techniques & Applications, Springer-Verlag, Telos, Santa Clara, CA

Wolfram, 1991
Wolfram, S., 1991, Mathematica, A System for Doing Mathematics by Computer, Second Edition, Addison-Wesley, Redwood City, CA

# Using Beta Distributions Efficiently in A Probabilistic Exposure Assessment

David E. Burmaster
Alceon Corporation
PO Box 382669, Cambridge, MA 02238-2669
tel: 617-864-4300; fax: 617-864-9954
deb@Alceon.com

## Abstract

In a probabilistic exposure assessment, an analyst must often develop a probability distribution to represent a random variable that has a fixed minimum and a fixed maximum. This manuscript shows how to fit a two-parameter Beta distribution to a sample data set, and then shows how to fit a constrained four-parameter Beta4 distribution to the same data -- thereby improving the fit and speeding the simulation by a factor of 5.

## 1.0 Introduction

In a probabilistic exposure assessment, an analyst must often develop a probability distribution to represent a random variable that has a fixed minimum and a fixed maximum. For example, a child may play in a park from 0 to 7 day/week. Similarly, the fraction of skin in contact with soils may range from 0 to 1, while the fraction of a nutrient or a toxicant absorbed in the gut may range from 0 to 1.

Beta distributions (of the first kind) (Mood et al, 1974; Evans et al, 1993) have several useful properties that an analyst may exploit in a probabilistic exposure assessment. First, a two-parameter Beta distribution can assume a wide variety of shapes, depending on the values taken by the two parameters. Second, a two-parameter Beta variate has a fixed minimum (zero) and a fixed maximum (one). Third, a two-parameter

Beta distribution can be scaled and translated, thereby creating the four-parameter Beta4 distribution (of the first kind). Finally, the method of Maximum Likelihood (Mood et al, 1974; Edwards, 1992) is a powerful way to fit either the two-parameter Beta distribution or the four-parameter Beta4 distribution to measured data. [EndNote 1].

Overall, the two-parameter Beta distribution and the four-parameter Beta4 distribution have many potential uses in probabilistic exposure assessments. This manuscript first shows how to fit a two-parameter Beta distribution to a sample data set, and then it shows how fit a four-parameter Beta4 distribution to the same data.

## 2.0 Data for the Sample Problem

To illustrate the techniques, we analyze some of the data reported in a recent manuscript (Magee et al, 1996; Table 1) in which the authors compiled 13 measurements of the absorption of polycyclic aromatic hydrocarbons from food in the guts of rats, hamsters, or humans. All the measurements fall between 0.7 and 1: 0.921, 0.89, 0.988, 0.887, 0.996, 0.967, 0.98, 0.87, 0.869, 0.94, 0.75, 0.97, and 0.938. In this manuscript, we accept these values as *bona fide* measurements of a phenomenon exhibiting considerable variability.

## 3.0 The Two-Parameter Beta Distribution

Theory: A Beta distribution (of the first kind) with two parameters, $a > 0$ and $b > 0$, describes the random variable $X$ over the support $0 \le x \le 1$: (Mood et al, 1974; Evans et al, 1993):

$$X \sim \text{Beta}[\, x \mid a, b \,] \quad ; \quad x \in [\, 0, 1 \,] \qquad \text{Eqn 1}$$

The random variable $\underline{X}$ has this probability density function (PDF):

$$f_X(x) = \frac{x^{a-1} \bullet (1 - x)^{b-1}}{\beta(a, b)} \qquad \text{Eqn 2}$$

where the denominator is the beta function, $\beta(a, b) = \int_0^1 u^{a-1} (1-u)^{b-1} du$ (Abramowitz & Stegun, 1964). The first two central moments of this distribution are, respectively, its expected value and its variance:

$$E[\underline{X}] = \frac{a}{a + b} \qquad \text{Eqn 3}$$

$$V[\underline{X}] = \frac{a \bullet b}{(a + b)^2 (a + b + 1)} \qquad \text{Eqn 4}$$

While the Beta distribution is extremely flexible over the support [ 0, 1 ] and finds many uses in statistics, two of its special properties are key in exposure assessment. First, when $a > 1$ and $b > 1$, the distribution has one mode:

$$\text{Mode}[\underline{X}] = \frac{a - 1}{a + b - 2} \qquad \text{Eqn 5}$$

Second, Beta variates are much easier and faster to simulate than when a and b are integers. [EndNote 2]. Further, the speed of the algorithm commonly used to simulate a Beta variate with integer parameters decreases as the sum (a + b) increases (Evans et al, 1993).

Exploratory Data Analysis: After exploratory data analysis to visualize the patterns in the data (Tukey, 1977; Chambers et al, 1983; Cleveland, 1985), we decided to fit a two-parameter Beta distribution to the data.

Application: We fit the Beta distribution to the data by maximizing the loglikelihood function (Edwards, 1992) in Mathematica™ (Wolfram, 1991; Wickham-Jones, 1994). In Figure 1, the solid line depicts the cumulative distribution function (CDF) for the best-fit distribution with noninteger parameters: $\underline{X}_1 \sim \text{Beta}[\, x \mid \hat{a} = 13.7035,\ \hat{b} = 1.17996 \,]$; and the dashed line depicts the CDF for the best-fit distribution with small integer parameters: $\underline{X}_2 \sim \text{Beta}[\, x \mid a = 14,\ b = 1 \,]$. The maxima of the loglikelihood functions for distributions $\underline{X}_1$ and $\underline{X}_2$ are 20.18 and 19.85, respectively. At this point, we have an excellent fit with noninteger parameters and an adequate fit with integer parameters.

## 4.0 The Four-Parameter Beta Distribution

Theory: A Beta4 distribution (of the first kind) with four parameters, $a > 0$, $b > 0$, $c > 0$, and $d > 0$ describes the random variable $\underline{Y}$ over the support $d \le y \le (c+d)$: (Mood et al, 1974; Evans et al, 1993)

$$\underline{Y} \quad \sim \quad \text{Beta4}[\, y \mid a, b, c, d \,] \quad ; \quad y \in [\, d, c + d \,] \qquad \text{Eqn 6}$$

This distribution arises from and can be simulated as a scaled and translated two-parameter Beta distribution:

$$\underline{Y} \quad = \quad c \bullet \underline{X} + d \quad ; \quad y \in [\, d, c + d \,] \qquad \text{Eqn 7}$$

The random variable $\underline{Y}$ has this PDF:

$$f_Y(y) \quad = \quad f_X\left(\frac{y - d}{c}\right) \qquad \text{Eqn 8}$$

and these central moments:

$$E[\underline{Y}] = c \bullet E[\underline{X}] \qquad \text{Eqn 9}$$

$$V[\underline{Y}] = c^2 \bullet V[\underline{X}] \qquad \text{Eqn 10}$$

Again, for $a > 1$ and $b > 1$, the Beta4 distribution has one mode:

$$\text{Mode}[\underline{Y}] = c \bullet \text{Mode}[\underline{X}] + d \qquad \text{Eqn 11}$$

Finally, when $c = 1 - d$ and $0 \leq d < 1$, the random variate $\underline{Y}$ has support [d, 1].

Exploratory Data Analysis: After further exploratory data analysis, we decided to fit a Beta4 distribution with a and b constrained to integers for speed in simulation.

Application: First, we fit the Beta4 distribution to the data by maximizing the loglikelihood function (Edwards, 1992) in Mathematica™ subject to the constraints: $a = 4$, $b = 1$, and $c = 1 - d$. In Figure 2, a solid line depicts the CDF for this fitted Beta4 distribution, $\underline{Y}_1 \sim \text{Beta4}[\, y \mid \hat{a} = 4, \hat{b} = 1, \hat{c} = 0.397303, \hat{d} = 0.602697 \,]$, plotted against the empirical CDF for the data. The maximum of the loglikelihood function for $\underline{Y}_1$ is 20.24, indicating that this constrained four-parameter Beta4 distribution fits the data better than the best two-parameter Beta distribution. For comparison, the dashed line in Figure 2 depicts the CDF for $\underline{X}_2$.

## 5.0 Discussion

Figure 3 shows the excellent correspondence between the best-fit two-parameter Beta distribution and the best-fit constrained four-parameter Beta4 distribution as a QQ-plot

(Chambers et al, 1983) for 100 realizations from $\underline{Y}_1$ and $\underline{X}_1$. In this figure, the diagonal line shows the locus of perfect match. In this application, it is inconsequential that $\underline{X}_1$ and $\underline{Y}_1$ diverge at the low extreme of the distribution (i.e., $\underline{X}_1$ will generate values smaller than d = 0.602697 less than one percent of the time but $\underline{Y}_1$ will never generate a value smaller than d).

While it is not surprising that an unconstrained four-parameter distribution can fit the data better than a two-parameter distribution, it is pleasing to note that $\underline{Y}_1$ (constrained, then optimized) fits the data better than $\underline{X}_1$ (unconstrained, optimized) as seen by comparing the maxima of the loglikelihood functions. Not only does $\underline{Y}_1$ fit the data better than $\underline{X}_1$, Mathematica™, for example, simulates $\underline{Y}_1$ some 5 times faster than it simulates $\underline{X}_1$. Further, $\underline{Y}_1$ has integer values for parameters a and b, so many popular Monte Carlo simulation programs that cannot simulate $\underline{X}_1$ can simulate $\underline{Y}_1$ .

## EndNotes

1. The analyst can use the "profile likelihood method" (Edwards, 1992) to find joint confidence regions for the parameters.
2. Most popular software packages accept only integer parameters.

## Acknowledgments

Brian H. Magee, Stephanie M. Vaughn, and Andrew M. Wilson provided many helpful comments during the preparation of this manuscript. Alceon Corporation funded this research.

## Trade Marks

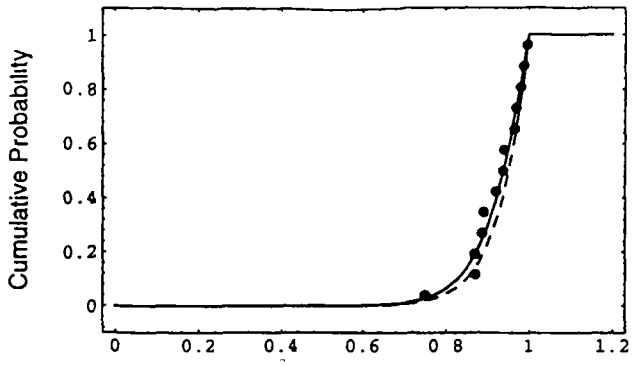



Figure 1: CDFs for $\underline{X}_1$ (solid) and $\underline{X}_2$ (dashed) plotted against the data

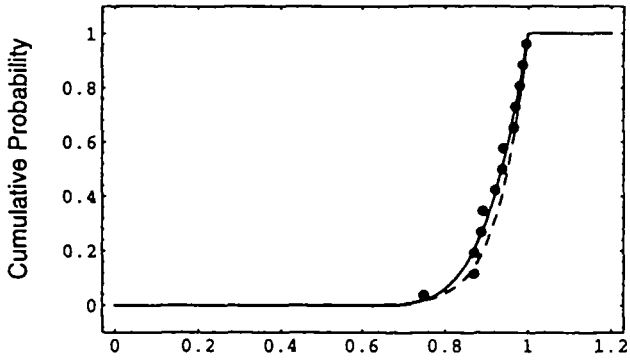



Figure 2: CDFs for $\underline{Y}_1$ (solid) and $\underline{X}_2$ (dashed) plotted against the data

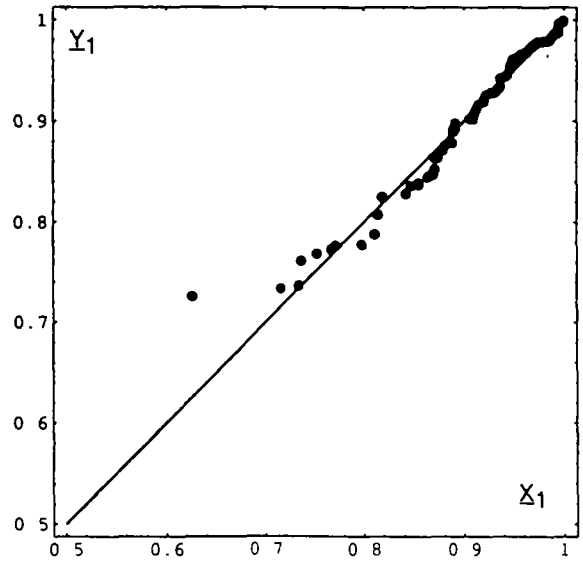



Figure 3: QQ-Plot for $\underline{Y}_1$ vs $\underline{X}_1$ (n = 100)

## References

Abramowitz & Stegun, 1964
Abramowitz, M. and I.A. Stegun, Eds, 1964, Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables, National Bureau of Standards, Applied Mathematics Series Number 55, Issued June 1964, Tenth Printing with corrections in December 1972, Superintendent of Documents, US Government Printing Office, Washington, DC

Chambers et al, 1983
Chambers, J.M., W.S. Cleveland, B. Kleiner, and P.A. Tukey, 1983, Graphical Methods for Data Analysis, Wadsworth International Group, Belmont, CA, and Duxbury Press, Boston, MA

Cleveland, 1985
Cleveland, W.S., 1985, The Elements of Graphing Data, Wadsworth Advanced Books, Monterey, CA

Edwards, 1992
Edwards, A.W.F., 1992, Likelihood, Expanded Edition, John Hopkins University Press, Baltimore, MD

Evans et al, 1993
Evans, M., N. Hastings, and J.B. Peacock, 1993, Statistical Distributions; Second Edition, John Wiley & Sons, Inc., New York, NY

Magee et al, 1996
Magee, B.H., P.D. Anderson, and D.E. Burmaster, 1996, Absorption Adjustment Factor Distributions (AAF) Distributions for Polycyclic Aromatic Hydrocarbons (PAHs), Human and Ecological Risk Assessment, in review

Mood et al, 1974
Mood, A.M., F.A. Graybill, and D.C. Boes, 1974, Introduction to the Theory of Statistics, Third Edition, McGraw Hill, New York, NY

Tukey, 1977
Tukey, J.W., 1977, Exploratory Data Analysis, Addison-Wesley, Reading, MA

Wickham-Jones, 1994
Wickham-Jones, T., 1994, Mathematica Graphics, Techniques & Applications, Springer-Verlag, Telos, Santa Clara, CA

Wolfram, 1991
Wolfram, S., 1991, Mathematica, A System for Doing Mathematics by Computer, Second Edition, Addison-Wesley, Redwood City, CA